

BIOMETRIKA

165

BIOMETRIKA

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

MANAGING EDITOR

E. S. PEARSON

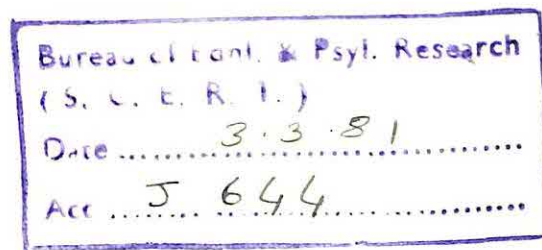
ASSOCIATE EDITOR

M. G. KENDALL



VOLUME 46

1959



8254/-

ISSUED BY

THE BIOMETRIKA OFFICE, UNIVERSITY COLLEGE, LONDON

PRINTED AT THE UNIVERSITY PRESS, CAMBRIDGE

PRINTED IN GREAT BRITAIN

CONTENTS OF VOLUME 46

Memoirs and Miscellanea

	PAGE
BANTEGUI, CELIA G. (<i>See</i> PILLAI, K. C. S.)	
BARRACLOUGH, ELIZABETH D. and PAGE, E. S. Tables for Wald tests for the mean of a normal distribution	169
BARTHOLOMEW, D. J. A test of homogeneity for ordered alternatives	36
BARTHOLOMEW, D. J. A test of homogeneity for ordered alternatives. II	328
BARTON, D. E. and DAVID, F. N. Sequential occupancy	218
BHATE, D. H. Approximation to the distribution of sample size for sequential tests. I. Tests of simple hypotheses	130
BIGGERS, J. D. The estimation of missing and mixed-up observations in several experimental designs	91
BOSE, R. C. and GUPTA, S. S. Moments of order statistics from a normal population	433
BOX, G. E. P. and LUCAS, H. L. Design of experiments in non-linear situations	77
BUSH, K. A. and OLKIN, I. Extrema of quadratic forms with applications to statistics	483
CLEMANS, K. G. Confidence limits in the case of the geometric distribution	260
CONOLLY, B. W. The busy period in relation to the queueing process GI/M/1	246
CROW, E. L. and GARDNER, R. S. Confidence intervals for the expectation of a Poisson variable	441
DARROCH, J. N. The multiple-recapture census II. Estimation when there is immigration or death	336
DARWIN, J. H. Note on a three-decision test for comparing two binomial populations	106
DARWIN, J. H. Note on the comparison of several realizations of a Markoff chain	412
DAVID, F. N. The z-test and symmetrically distributed random variables	123
DAVID, F. N. (<i>See</i> BARTON, D. E.)	
DAVID, H. A. Tournaments and paired comparisons	139
DURBIN, J. Efficient estimation of parameters in moving-average models	306
DURBIN, J. A note on the application of Quenouille's method of bias reduction to the estimation of ratios	477
VAN ELTEREN, PH. and NOETHER, G. E. The asymptotic efficiency of the χ^2 -test for a balanced incomplete block design	475
GABRIEL, K. R. The distribution of the number of successes in a sequence of dependent trials	454
GARDNER, R. S. (<i>See</i> CROW, E. L.)	
GOODMAN, L. A. Partial tests for partial taus	425
GUPTA, S. S. (<i>See</i> BOSE, R. C.)	
HAIGHT, F. A. Overflow at a traffic light	420
HARTLEY, H. O. The efficiency of internal regression for the fitting of the exponential regression	293
HEATHCOTE, C. R. and MOYAL, J. E. The random walk (in continuous time) and its application to the theory of queues	400

	PAGE
JOHNSON, N. L. On an extension of the connexion between Poisson and χ^2 distributions	352
JOWETT, G. H. and WRIGHT, WENDY, M. Jump analysis	386
KHATRI, C. G. On certain properties of power-series distributions	486
KIMBALL, A. W. and LEACH, E. Approximate linearization of the incomplete β -function	214
LAWLEY, D. N. Tests of significance in canonical analysis	59
LEACH, E. (<i>See</i> KIMBALL, A. W.)	
LESLIE, D. C. M. Determination of parameters in the Johnson system of probability distributions	229
LEVINE, J. Monomial-monomial symmetric function tables	205
LIPTON, S. (<i>See</i> PATTERSON, H. D.)	
LOYD, D. E. Note on a problem of estimation	231
LUCAS, H. L. (<i>See</i> BOX, G. E. P.)	
MALLOWS, C. L. On the probability integral transformation	481
MERCER, A. and SMITH, C. S. A random walk in which the steps occur randomly in time	30
MILES, R. E. The complete amalgamation into blocks, by weighted means, of a finite set of real numbers	317
MITTON, R. G. and MORGAN, F. R. The design of factorial experiments: a survey of some schemes requiring not more than 256 treatment combinations	251
MORGAN, F. R. (<i>See</i> MITTON, R. G.)	
MOYAL, J. E. (<i>See</i> HEATHCOTE, C. R.)	
NASS, C. A. G. The χ^2 test for small expectations in contingency tables, with special reference to accidents and absenteeism	365
NOETHER, G. E. (<i>See</i> VAN ELTEREN, PH.)	
OLKIN, I. (<i>See</i> BUSH, K. A.)	
PACHARES, J. Table of the upper 10 % points of the 'studentized' range	461
PAGE, E. S. (<i>See</i> BARRACLOUGH, ELIZABETH D.)	
PATTERSON, H. D. and LIPTON, S. An investigation of Hartley's method for fitting an exponential curve	281
PEARSON, E. S. Note on Mr Quenouille's Edgeworth Type A transformation	203
PEARSON, E. S. Note on an approximation to the distribution of non-central χ^2	364
PHILLIPS, A. W. The estimation of parameters in systems of stochastic differential equations	67
PILLAI, K. C. S. Upper percentage points of the extreme studentized deviate from the sample mean	473
PILLAI, K. C. S. and BANTEGUI, CELIA G. On the distribution of the largest of six roots of a matrix in multivariate analysis	237
PILLAI, K. C. S. and SAMSON, P., Jr. On Hotelling's generalization T^2	160
PILLAI, K. C. S. and TIENZO, B. P. On the distribution of the extreme studentized deviate from the sample mean	467
QUENOUILLE, M. H. Tables of random observations from standard distributions	178
RAMASUBBAN, T. A. The generalized mean differences of the binomial and Poisson distributions	223

	PAGE
RAO, C. R. Some problems involving linear hypotheses in multivariate analysis	49
SAMSON, P., Jr. (See PILLAI, K. C. S.)	
SANKARAN, M. On the non-central χ^2 distribution	235
SAW, J. G. Estimation of the normal population parameters given a singly censored sample	150
SHENTON, L. R. The distribution of moment estimators	296
SMITH, C. S. (See MERCER, A.)	
SMITH, W. L. On the cumulants of renewal processes	1
SOMERS, R. H. The rank analogue of product-moment partial correlation and regression, with application to manifold, ordered contingency tables	241
SRIVASTAVA, A. B. L. Effect of non-normality on the power of the analysis of variance test	114
TIENZO, B. P. (See PILLAI, K. C. S.)	
WRIGHT, WENDY M. (See JOWETT, G. H.)	

Book Reviews

ANDERSON, O. <i>Probleme der Statistischen Methodenlehre in den Sozialwissenschaften</i> (Third edition)	C. A. MOSER	272
BERS, L. <i>Surveys in Applied Mathematics</i> . Vol. 3. <i>Mathematical Aspects of Subsonic and Transonic Gas Dynamics</i>	C. L. MALLOWS	499
BEVERTON, R. J. H. and HOLT, S. J. <i>On the Dynamics of Exploited Fish Populations</i>	F. N. DAVID	495
BLISS, C. I. <i>Periodic Regression in Biology and Climatology</i>	F. N. DAVID	495
BOOLE, G. <i>Calculus of Finite Differences</i>	F. N. DAVID	271
BURINGTON, R. S. and MAY, D. C. <i>Handbook of Probability and Statistics with Tables</i>	F. N. DAVID	497
CHURCHMAN, C. W., ACKOFF, R. L. and ARNOFF, E. L. <i>Introduction to Operations Research</i>	ALISON DOIG	274
CLEW, V. <i>Experimental Designs in Industry</i>	N. L. JOHNSON	266
COX, D. R. <i>Planning of Experiments</i>	C. L. MALLOWS	492
<i>Demografía</i> , Vol. 1, Part 1	N. H. CARRIER	272
DUBOIS, P. H. <i>Multivariate Correlational Analysis</i>	J. DURBIN	268
FISZ, M. <i>Wahrscheinlichkeitsrechnung und Mathematische Statistik</i>	REGINA C. ELANDT	494
FORSYTHE, G. E. and ROSENBLOOM, P. C. <i>Surveys in Applied Mathematics</i> . Vol. 5. <i>Numerical Analysis and Partial Differential Equations</i>	C. L. MALLOWS	499
GASS, S. I. <i>Linear Programming: Methods and Applications</i>	S. VAJDA	494
GOODIER, J. N. and HODGE, P. G. <i>Surveys in Applied Mathematics</i> . Vol. 1. <i>Elasticity and Plasticity</i>	C. L. MALLOWS	499
GOODMAN, R. <i>Teach yourself Statistics</i>	D. E. BARTON	269
GUPTA, H., GWYTHYR, C. E. and MILLER, J. C. P. <i>Tables of Partitions</i>	D. E. BARTON	498
HOHN, F. E. <i>Elementary Matrix Algebra</i>	C. L. MALLOWS	497

	PAGE
HUTCHINSON, SIR JOSEPH. <i>Genetics and Improvement of Tropical Crops</i>	A. R. G. OWEN 273
KAPLANSKY, I. <i>et al.</i> <i>Surveys in Applied Mathematics</i> . Vol. 4. <i>Some Aspects of Analysis and Probability</i>	C. L. MALLOWS 499
KENDALL, M. G. and STUART, A. <i>The Advanced Theory of Statistics</i> . Vol. 1. <i>Distribution Theory</i>	E. S. PEARSON 491
LEIMANIS, E. and MINORSKY, N. <i>Surveys in Applied Mathematics</i> . Vol. 2. <i>Dynamics and Nonlinear Mechanics</i>	C. L. MALLOWS 499
LI, J. C. R. <i>Introduction to Statistical Inference</i>	F. N. DAVID 269
MOORE, P. G. <i>Principles of Statistical Techniques. A First Course from the beginnings for Schools and Universities</i>	A. P. PENFOLD 276
MORSE, P. M. <i>Queues, Inventories and Maintenance</i>	D. M. G. WISHART 496
OLSON, E. C. and MILLER, R. L. <i>Morphological Integration</i>	F. N. DAVID 273
OSGOOD, W. F. <i>Functions of Real and Complex Variables</i>	T. ESTERMANN 270
PANIZZON, G. <i>Contribuito allo Studio delle Tavole di Nuzialità</i>	C. A. B. SMITH 495
PECK, L. G. and HAZELWOOD, R. N. <i>Finite Queueing Tables</i>	D. M. G. WISHART 494
PILLAI, K. C. S. <i>Concise Tables for Statisticians</i>	D. E. BARTON 270
POPPER, K. R. <i>The Logic of Scientific Discovery</i>	M. G. KENDALL 265
QUENOUILLE, M. H. <i>Analysis of Multiple Time-Series</i>	G. M. JENKINS 267
QUENOUILLE, M. H. <i>Fundamentals of Statistical Reasoning</i>	C. L. MALLOWS 268
RIORDAN, J. <i>An Introduction to Combinatorial Analysis</i>	F. N. DAVID 271
SALMON, G. <i>A Treatise on the Analytic Geometry of Three Dimensions</i> (Seventh edition) Vol. 1, ...	F. N. DAVID 271
SALZER, H. E. <i>et al.</i> <i>Table for the Solution of Cubic Equations</i>	D. E. BARTON 499
<i>Statistical Research Monographs</i> , Note regarding	264
TAYLOR, A. E. <i>Introduction to Functional Analysis</i>	R. E. EDWARDS 497
TORGERSON, W. S. <i>Theory and Methods of Scaling</i>	C. BURT 493
VAZSONYI, A. <i>Scientific Programming in Business and Industry</i>	S. VAJDA 274
WINNICK, L. <i>American Housing and its Uses</i>	MARIAN BOWLEY 273

BIOMETRIKA PUBLICATIONS

*Issued by the Cambridge University Press, Bentley House, London, N.W. 1
and obtainable from any bookseller*

Tables of the Incomplete B-Function

59 pages of Introduction and 494 pages of Tables

EDITED BY KARL PEARSON

Price: 55s. net

Tables of the Incomplete Γ -Function

31 pages of Introduction and 164 pages of Tables

EDITED BY KARL PEARSON

Price: 42s. net

Tables of the Complete and Incomplete Elliptic Integrals

(from LEGENDRE'S *Traité des Fonctions Elliptiques*. With autographed portrait of LEGENDRE)

39 pages of Introduction by KARL PEARSON and 94 pages of Tables

Price: 12s. 6d. net

Tables of the Ordinates and Probability Integral of the Distribution of the Correlation Coefficient in Small Samples

BY F. N. DAVID

38 pages of Introduction, 55 pages of Tables, 10 Diagrams and 4 Charts

Price: 17s. 6d. net

Biometrika Tables for Statisticians, Vol. I

EDITED BY E. S. PEARSON and H. O. HARTLEY for the Biometrika Trust

102 pages of Introduction and 136 pages of Tables

Price: 25s. net

The Life, Letters and Labours of Francis Galton, Vols. I, II, IIIA, & IIIB

BY KARL PEARSON, F.R.S.

Price: £3. 3s. net

Karl Pearson: An Appreciation of Some Aspects of his Life and Work

BY E. S. PEARSON

Price: 15s. net

A Bibliography of the Statistical and Other Writings of Karl Pearson

COMPILED BY G. M. MORANT, with the assistance of B. L. WELCH

Price: 6s. net

"Student's" Collected Papers

EDITED BY E. S. PEARSON and

JOHN WISHART with a FOREWORD BY LAUNCE MCMULLEN

Price: 21s. net

Karl Pearson's Early Statistical Papers

Reprinted by photo-lithography for the Biometrika Trust, with the permission of the original publishers. The Volume contains eleven papers, including the more important of the memoirs entitled "Mathematical Contributions to the Theory of Evolution", first published in the *Philosophical Transactions* of the Royal Society. The original paper deriving the χ^2 -distribution, published in 1900 in the *Philosophical Magazine*, is also included.

Price: 25s. net

NEW STATISTICAL TABLES: SEPARATES RE-ISSUED FROM BIOMETRIKA

To be obtained from
BIOMETRIKA OFFICE, UNIVERSITY COLLEGE, LONDON, W.C.1

- I. From *Biometrika*, Vols. 22, 27 and 28
Tests of Normality. By E. S. PEARSON and R. C. GEARY
Price 2s. 6d., post free
- II. From *Biometrika*, Vol. 32, pp. 168-181 and 188-189
(1) Table of percentage points of the incomplete beta-function
(2) Table of percentage points of the χ^2 distribution
Stitched together with introductory matter. Price 2s. 6d., post free
- III. From *Biometrika*, Vol. 32, pp. 300-310
(1) Table of the probability integral of the range in samples from a normal population
(2) Table of the percentage points of the range
(3) Table of the percentage points of the *t*-distribution
Stitched together with introductory matter. Price 2s. 6d., post free
- IV. From *Biometrika*, Vol. 33, pp. 73-88
Table of percentage points of the inverted beta (F) distribution
With introductory matter. Price 2s. 6d., post free
- V. From *Biometrika*, Vol. 33, pp. 252-265
(1) Table of the probability integral of the mean deviation in samples from a normal population
(2) Table of the percentage points of the mean deviation
Stitched together with introductory matter. Price 2s. 6d., post free
- VI. From *Biometrika*, Vol. 33, pp. 296-304
Table for testing the homogeneity of a set of estimated variances
With introductory matter. Price 2s., post free
- VII. From *Biometrika*, Vol. 35, pp. 145-156
Table of significance levels for the Fisher-Yates test of significance in 2×2 contingency tables. By D. J. FINNEY
With introductory matter. Price 2s. 6d., post free
- VIII. From *Biometrika*, Vol. 35, pp. 191-201
Table for the calculation of working probits and weights in probit analysis. By D. J. FINNEY and W. L. STEVENS
With introductory matter. Price 2s. 6d., post free
- IX. From *Biometrika*, Vol. 36, pp. 267-289
Tables of autoregressive series. By M. G. KENDALL
With introductory matter. Price 2s. 6d., post free
- X, XIV, XVIII, and XX. From *Biometrika*, Part 1 from Vol. 36, pp. 431-449, Parts 2 and 3 from Vol. 38, pp. 435-462, Part 4 from Vol. 40, pp. 427-446, and Part 5 from Vol. 42, pp. 223-242
Tables of symmetric functions. By F. N. DAVID and M. G. KENDALL
With introductory matter. Price 14s. 6d., post free
(Part 1, 2s. 6d.; Parts 2 and 3, 4s.; Part 4, 4s.; Part 5, 4s.)
- XI. From *Biometrika*, Vol. 39, p. 190 and Vol. 43, pp. 449-451
Tables of percentage points of the extreme "Studentized" deviate from the sample mean. By K. R. NAIR and H. A. DAVID
With introductory matter. Price 1s., post free
- XII. From *Biometrika*, Vol. 37, pp. 168-172 and pp. 313-325
(1) Table of the probability integral of the *t*-distribution
(2) Table of the χ^2 integral, and of the cumulative Poisson distribution. By H. O. HARTLEY and E. S. PEARSON
Stitched together with introductory matter. Price 5s., post free

NEW STATISTICAL TABLES: continued

XIII. From *Biometrika*, Vol. 38, pp. 112-130

Charts of the power function for analysis of variance tests, derived from the non-central F-distribution. By E. S. PEARSON and H. O. HARTLEY

With introductory matter. Price 2s. 6d., post free

XV. From *Biometrika*, Vol. 38, pp. 423-426

A chart for the incomplete beta-function and the cumulative binomial distribution. By H. O. HARTLEY and E. R. FITCH

With introductory matter and ruler scale. Price 2s. 6d., post free

XVI. From *Biometrika*, Vol. 40, pp. 70-73

Tables of the angular transformation. By W. L. STEVENS

With introductory matter. Price 1s., post free

XVII. From *Biometrika*, Vol. 40, pp. 74-86

Tests of significance in a 2×2 contingency table: extension of Finney's table (No. VII).

Computed by R. LATSCHA

With introductory matter. Price 2s. 6d., post free

XIX. From *Biometrika*, Vol. 41, pp. 253-260

Tables of generalized k-statistics. By S. H. ABDEL-ATY

With introductory matter. Price 2s., post free

XXI. From *Biometrika*, Vol. 42, pp. 494-511

A new form of table for significance tests in a 2×2 contingency table. By P. ARMSEN

With introductory matter. Price 2s. 6d., post free

XXII. From *Biometrika*, Vol. 43, pp. 388-403

Tables for certain applications of sequential methods in the analysis of variance. By W. D. RAY

With introductory matter. Price 2s. 6d., post free

XXIII. From *Biometrika*, Vol. 43, pp. 423-435

Table for determining confidence limits for a proportion in binomial sampling. By EDWIN L. CROW

With introductory matter. Price 2s. 6d., post free

XXIV. From *Biometrika*, Vol. 44, pp. 411-419

Tables for estimating the normal distribution function of normit analysis. Part I. Tables and description of their use. By JOSEPH BERKSON

With introductory matter. Price 2s. 6d., post free

XXV. From *Biometrika*, Vol. 44, pp. 482-489

Table of significance points for a two-sample t-test based on range. By P. G. MOORE

With introductory matter. Price 2s. 6d., post free

XXVI. From *Biometrika*, Vols. 44 & 45

Tables of the upper percentage points of the generalized beta distribution. By F. G. FOSTER and D. H. REES

With introductory matter. Price 5s., post free

From *Biometrika*, Vol. 44, pp. 490-514

A bibliography on the theory of queues. By ALISON DOIG

Price 5s., post free

From *Biometrika*, Vol. 45, pp. 293-315

THOMAS BAYES'S Essay towards solving a problem in the doctrine of chances. [Reproduced from *Phil. Trans. Roy. Soc.* 1763, 53, 370-418.] With a biographical note by G. A. BARNARD.

Price 5s., post free

From *Biometrika*, Vol. 45, pp. 521-543

A bibliography on life testing and related topics. By WILLIAM MENDENHALL

Price 5s., post free

BIOMETRIKA INDEX. Comprising Subject Index for Vols. 1-37 and Author Index for Vols. 1-40, with Author Index Supplement covering Vols. 41-43.

Price 6s. or \$1.00, post free

STATISTICAL EXERCISES

Issued by the DEPARTMENT OF STATISTICS
UNIVERSITY COLLEGE, LONDON, W.C.1

Part I. Elementary Statistical Exercises.

Compiled by F. N. DAVID (91 pp.)

These exercises, published in 1953, were collected in connection with the numerical classwork undertaken by students during the first year of the B.Sc. Special Degree course in Statistics at University College. They deal with applications of the more elementary univariate and bivariate theory.

Price: 6s. 6d.

Part II. Statistical Exercises. Analysis of variance and associated techniques. (107 pp.)

Compiled by N. L. JOHNSON

This volume includes over 100 exercises on the following topics:

Analysis of variance techniques for the following experimental designs: randomized block, Latin square, confounding, split plots, fractional replication. *Linear and curvilinear regression. Dosage-mortality techniques. Discriminant functions. Time series. Curve fitting.*

Price 12s.

[A supplementary volume containing suggested solutions to a substantial proportion of the exercises in Part II is in preparation.]

Both Parts are reproduced by offset-litho processes from typescript and bound in stiff paper covers. Copies may be obtained from a bookseller or direct from the Department of Statistics, University College, Gower Street, London, W.C. 1. If ordered direct, payment must be made in advance and cheques made payable to University College London.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

VOL. 54, No. 285

CONTENTS

MARCH 1959

Maps Based on Probabilities. M. CHOYNOWSKI—Statistical Data Available for Economic Research on Certain Types of Recreation. MARION CLAWSON—Optimum Univariate Stratification. TORE DALENIUS and JOSEPH L. HODGES, Jr.—Some Leading Soviet Statistical Books of 1957. EBERHARD FELS—Measures of Association for Cross-Classifications: Further Discussion and References. LEO A. GOODMAN and WILLIAM H. KRUSKAL—Inhalation in Relation to Type and Amount of Smoking. E. CUYLER HAMMOND—Statisticians: Today and Tomorrow. WALTER E. HOADLEY—How Many of a Group of Random Numbers Will be Usable in Selecting a Particular Sample? HOWARD L. JONES—Testing the Hypotheses that a Binomial Probability Equals $1/2$. WILLIAM J. MACKINNON—The Fitting of Straight Lines when both Variables are Subject to Error. ALBERT MADANSKY—Some Problems of the Household Interview Design for the National Health Survey. THEODORE D. WOOLSEY and HAROLD NISSELSON—A Production Model and Continuous Sampling Plan. I. RICHARD SAVAGE—A Single Sampling Plan for Correlated Variables with a Single-Sided Specification Limit. K. C. SEAL—Panel Mortality and Panel Bias. MARION G. SOBOL—Publications Decisions and their Possible Effects on Inferences Drawn from Tests of Significance, or vice versa. THEODORE D. STERLING—Minimum Risk Specification Limits. FRED H. TINGEY and J. A. MERRILL—Linear Programming Techniques for Regression Analysis. HARVEY M. WAGNER—Simplified Beta-Approximations to the Kruskal-Wallis H. Test. DAVID L. WALLACE—Comments on 'The Simplest Signed-Rank Tests'. JOHN E. WALSH—Lower Bound Formulas for the Mean Inter-correlation Coefficient. RICHARD H. WILLIS—Book Reviews—Notes About Authors—Publications Received.

For information about the Association and membership application forms, write

THE SECRETARY, AMERICAN STATISTICAL ASSOCIATION,
1757 K. ST., N.W., WASHINGTON 6, D.C., U.S.A.

BIOMETRICS

Journal of the Biometric Society

VOL. 15, No. 1

TABLE OF CONTENTS

MARCH 1959

Host Variability in Dilution Experiments. PETER ARMITAGE—Restricted Selection Indices. OSCAR KEMPTHORNE and ARNE W. NORDSKOG—Equilibria in Auto-Tetraploids Under Natural Selection for a Simplified Model of Viabilities. P. A. PARSONS—The Analysis of a Non-Replicated Experiment Involving a Single Four-Course Rotation of Crops. H. D. PATTERSON—The Analysis of a Two-Phase Experiment. R. N. CURNOW—Analysis of Quadruple Rectangular Lattice Designs. JOHN LEROY FOLKS—An Examination of Some Methods of Comparing Several Rates or Proportions. MINDEL C. SHEPS—The Analysis of Experiments on Growth Rate. F. B. LEECH and M. J. R. HEALY—Calculation of Chi-Square to Test the No Three-Factor Interaction Hypothesis. MARVIN A. KASTENBAUM and DONALD E. LAMPHIEAR—Extra-Period Change-Over Designs. H. D. PATTERSON and H. L. LUCAS—Query: Replication of Non-Center Points in the Rotatable and Near-Rotatable Central Composite Design. G. E. P. BOX—Query: Significance of Difference Between Two Non-Independent Correlation Coefficients. E. J. WILLIAMS.

VOL. 15, No. 2

JUNE 1959

A Simple Method for Constructing Orthogonal Polynomials When the Independent Variable Is Unequally Spaced. D. S. ROBSON—The Estimation of Environmental and Genetic Trends from Records Subject to Culling. C. R. HENDERSON, OSCAR KEMPTHORNE, S. R. SEARLE and C. M. VON KROSIGK—Experimental Design in the Evaluation of Genetic Parameters. ALAN ROBERTSON—A Distribution-Free Asymptotic Method of Estimating, Testing, and Setting Confidence Limits for Heritability. LORRAINE SCHWARTZ and STANLEY WEARDEN—The Regression Analysis of Causal Paths. MALCOLM E. TURNER and CHARLES D. STEVENS—A Class of Two Replicate Incomplete Block Designs. J. ROY—The Centric Systematic Area-Sample Treated as a Random Sample. A. MILNE—Sensory Item Sorting. N. T. GRIDGEMAN—Orthogonal Contrasts in Slope Ratio Investigations. P. J. CLARINGBOLD—Note: A Confidence Interval on the Abscissa of the Point of Intersection of Two Fitted Linear Regressions. MARVIN A. KASTENBAUM—Note: On the Estimation of the Mean of Poisson Distribution from a Sample with the Zero Class Missing. J. O. IRWIN—Query: Differential Regression. JOHN T. WEBSTER.

Biometrics is published quarterly. The annual non-member subscription rate is £7.00. Subscriptions, inquiries, and orders for back issues should be sent to:

BIOMETRICS

DEPARTMENT OF STATISTICS, VIRGINIA POLYTECHNIC INSTITUTE
BLACKSBURG, VIRGINIA, U.S.A.

TRABAJOS DE ESTADISTICA

REVIEW PUBLISHED BY *INSTITUTO DE INVESTIGACIONES ESTADÍSTICAS*
OF THE *CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS*
MADRID, SPAIN

Vol. IX.

Cuad. III

CONTENTS

P. ZOROA—Convolución de histogramas.

D. MARAVALL—La adición de vectores aleatorios isótropos en un espacio de N dimensiones.

NOTAS.

E. CANSADO—Sobre la inversión de matrices de Leontief.

J. M. GARCIA GARCIA—Problema de la selección de la Cartera.

CRONICA.

BIBLIOGRAFIA.

CUESTIONES Y EJERCICIOS.

For everything in connection with works, exchanges and subscription write to Professor Sixto Rios, Instituto de Investigaciones Estadísticas, Consejo Superior de Investigaciones Científicas, Serrano no. 123, Madrid, Spain.

The Review is composed of three fascicles published three times a year (about 350 pp.), and its annual price is 100 pesetas for Spain and \$4.00 U.S.A. for all other countries.

Annals of Human Genetics

Formerly ANNALS OF EUGENICS

Edited by L. S. PENROSE

VOL. 23, Pt. 2

CONTENTS

FEBRUARY 1959

Testing association of metric traits and marker genes. DOROTHY C. LOWRY and F. T. SCHULTZ—Microcephaly in the Netherlands: a clinical and genetic study. J. VAN DEN BOSCH—Enzymatic evidence of a galactosemic trait of galactosemic children. H. N. KIRKMAN and E. BYNUM—Muscular dystrophy in childhood—the genetic aspect. HELEN BLYTH and R. PUGH—Distribution of birth weights of Chinese and Indian infants born in Singapore: birth weight as an index of maturity. JEAN MILLIS—Frequency of twinning in poor Chinese in the maternity hospital, Singapore. JEAN MILLIS—Frequencies of haptoglobin types in five populations. H. ELTON SUTTON, J. V. NEEL, F. B. LIVINGSTONE, G. BINSON, P. KUNSTADTER and L. E. TROMBLEY—Reviews.

VOL. 23, Pt. 3

MAY 1959

On the effect of maternal age and parity on birthweight of offspring (Indian data). N. K. NAMBOODIRI and V. BALAKRISHNAN—Non-random mating, and its effect on the rate of approach to homozygosity. G. A. WATTERSON—A new genetic population model, and its approach to homozygosity. G. A. WATTERSON—Some comments on Schützenberger's analysis of data on the human sex ratio. A. W. F. EDWARDS—Some statistical data on atypical cholinesterase of human serum. W. KALOW and D. R. GUNN—The numbers of aunts and uncles of normal and congenitally abnormal children. G. KNOX—Taste sensitivity to phenyl-thio-urea in tuberculosis and diabetes mellitus. O. ÅKESSON—The inheritance of celiac disease. C. CARTER, W. SHELDON and C. WALKER—The incidence of anencephaly in a polytypic population. A. G. SEARLE—Birth weight in cystic fibrosis of the pancreas. D. Y.-Y. HSIA—Familial Alzheimer's disease. LORNA WHEELAN—A note on the effects of method of ascertainment of segregation ratios. C. A. B. SMITH—Review.

*Subscription price 100s. net per volume of four quarterly parts (in U.S.A. \$17.50) post free.
Single issues, 30s. (in U.S.A. \$5.00) postage extra.*

CAMBRIDGE UNIVERSITY PRESS

BENTLEY HOUSE, 200 EUSTON ROAD, LONDON, N.W. 1

The Annals of Mathematical Statistics

The Official Journal of the Institute of Mathematical Statistics

VOL. 30, NO. 1

CONTENTS

MARCH, 1959

Invariance Theory and a Modified Minimax Principle. OSCAR WESLER—On Linear Associative Algebras Corresponding to Association Schemes of Partially Balanced Designs. R. C. BOSE and DALE M. MESNER—On a Characterization of the Triangular Association Scheme. S. S. SHRIKHANDE—On a Generalisation of the Kronecker Product Designs. B. V. SHAH—Application of a Measure of Information to the Design and Comparison of Regression Experiments. M. STONE—Note on Estimating Information. COLIN R. BLYTH—Unbiased Sequential Estimation for Binomial Populations. MORRIS H. DEGROOT—A Single-Sample Multiple-Decision Procedure for Selecting the Multinomial Event Which Has the Highest Probability. ROBERT E. BECHHOFFER, SALAH ELMAHRY and NORMAN MORSE—A Property of the Multinomial Distribution. HARRY KESTEN and NORMAN MORSE—A Classification Problem Involving Multinomials. OSCAR WESLER—A Problem in Optimum Filtering with Finite Data. GOPINATH KALLIANPUR—On the Limiting Distribution of the Number of Coincidences Concerning Telephone Traffic. L. TAKÁCS—On the Integrodifferential Equation of Takács. II. EDGAR REICH—Symmetrizable Markov Matrices. H. P. KRAMER—On Some Statistical Tests for M th Order Markov Chains. LEO A. GOODMAN—Consensus of Subjective Probabilities: The Pari-Mutuel Method. EDMUND EISENBERG and DAVID GALE—A Note on Perfect Probability. GOPINATH KALLIANPUR—On the Distribution of the Kolmogorov-Smirnov D -Statistic. PEDRO EGYDIO DE OLIVEIRA CARVALHO—Equality of More Than Two Variances and of More Than Two Dispersion Matrices Against Certain Alternatives. R. GNANADESIKAN—On the Theory of BAN Estimates. ROBERT A. WIJSMAN—Estimation of the Medians for Dependent Variables. OLIVE JEAN DUNN—A Consistent Estimator of a Component of a Convolution. WILLIAM R. GAFFEY—Bayes and Minimax Procedures in Sampling from Finite and Infinite Populations. OM P. AGGARWAL—An Approximation Useful in Univariate Stratification. GUNNAR EKMAN—Truncation and Tests of Hypotheses. OM P. AGGARWAL and IRWIN GUTTMAN—NOTES: Bartlett Decomposition and Wishart Distribution. A. M. KSHIRSAGAR—Some Bounds on the Distribution Functions of the Largest and Smallest Roots of Normal Determinantal Equations. RAY MICKEY—Etc., etc.—ABSTRACTS OF PAPERS—NEWS AND NOTICES—REPORT OF THE MONTEREY, CALIFORNIA, MEETING—FINAL EDITORIAL REPORT—PUBLICATIONS RECEIVED.

Subscription rate \$12.00 per year in the United States and Canada and \$10.00 per year elsewhere

ADDRESS ORDERS FOR SUBSCRIPTIONS AND BACK NUMBERS TO
PROFESSOR A. H. BOWKER, TREASURER, INSTITUTE OF MATHEMATICAL
STATISTICS, DEPARTMENT OF STATISTICS, STANFORD UNIVERSITY,
STANFORD, CALIFORNIA, U.S.A.

ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

Contents of Vol. 27, No. 1, January 1959, include:

- C. J. VAN EIJK and J. SANDEE. Quantitative Determination of an Optimum Economic Policy.
HENRI THEIL. The Aggregation Implications of Identifiable Structural Macrorelations.
ROBERT SOLOW. Competitive Valuation in a Dynamic Input-Output System.
LIONEL W. MCKENZIE. On the Existence of General Equilibrium for a Competitive Market.
R. L. BASMANN. The Computation of Generalized Classical Estimates of Coefficients in a Structural Equation.
KENNETH J. ARROW, H. D. BLOCK and L. HURWICZ. On the Stability of Competitive Equilibrium.
H. F. LYDALL. The Distribution of Employment Incomes.
MURRAY BROWN. On the Structure of Stochastic Difference Equation Models.
ROBERT SUMMERS. A Note on Least Squares Bias in Household Expenditure Analysis.
S. J. PRAIS. A Comment.
BOOK REVIEWS.

Published Quarterly

Subscription rate available on request

The Econometric Society is an international society for the advancement of economic theory in its relation to statistics and mathematics.

Subscriptions to *Econometrica* and inquiries about the work of the Society and the procedure in applying for membership should be addressed to

RICHARD RUGGLES, *Secretary*

THE ECONOMETRIC SOCIETY, BOX 1264, YALE UNIVERSITY
NEW HAVEN, CONNECTICUT, U.S.A.

SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

EDITED BY P. C. MAHALANOBIS

Vol. 20, parts 3 and 4, 1958

CONTENTS

- The Scope of Biological Statistics. J. B. S. HALDANE.
Examples of Inconsistency of Maximum Likelihood Estimates. R. R. BAHADUR.
Maximum Likelihood Estimation for the Multinomial Distribution with Infinite Number of Cells. C. RADHAKRISHNA RAO.
A Remark on Strong Measurability. V. S. VARADARAJAN.
A Useful Convergence Theorem. V. S. VARADARAJAN.
On Statistics Independent of Sufficient Statistics. D. BASU.
On Configurations and Non-Isomorphism of Some Incomplete Block Designs. M. ATIQULLAH.
Tables of Random Normal Deviates. J. M. SENGUPTA and NIKHILESH BHATTACHARYA.
On Sampling With and Without Replacement. D. BASU.
Diffusion by Discrete Movements. H. C. GUPTA.
On the Distribution of the Maximum Value of an Equally Correlated Sample from a Normal Population. A. KUDÔ.
Hypergeometric Expansions for Incomplete Moments of the Bivariate Normal Distribution. A. R. KAMAT.
Incomplete Moments of the Trivariate Normal Distribution. A. R. KAMAT.
A Note About a Family of Perks' Distribution. JOSEPH TALACKO.
Recent Experiments in Statistical Sampling in the Indian Statistical Institute. P. C. MAHALANOBIS.

SUBSCRIPTION

CURRENT

BACK NUMBERS

	<i>per volume</i>	<i>per issue</i>	<i>per volume</i>	<i>per issue</i>
INDIA	Rs. 30	Rs. 10	Rs. 45	Rs. 12/8
FOREIGN	\$10.00	\$3.50	\$15.00	\$4.50

Subscriptions and orders for back numbers should be sent to

Statistical Publishing Society, 204/1 Barrackpore Trunk Road, Calcutta - 35

ROYAL STATISTICAL SOCIETY

THE JOURNAL OF THE ROYAL STATISTICAL SOCIETY is published in two series: SERIES A (GENERAL), four issues a year, annual subscription £4. 4s. post free; SERIES B (METHODOLOGICAL), two issues a year, annual subscription £3. 2s. post free.

SERIES A (GENERAL), VOL. 122, PART 1, 1959

The Long-term Trend in the Size Distribution of Income. By H. F. LYDALL (with Discussion)—Report of the Committee on the Supply of and Demand for Statisticians—Introduced by Professor E. S. PEARSON (with Discussion)—Size of Company and Other Factors in Dividend Policy. By P. SARGANT FLORENCE—Reviews of Books, Statistical and Current Notes, Additions to Library.

SERIES B (METHODOLOGICAL), VOL. 21, PART 1, 1959

Geometric Distributions in the Theory of Queues. By C. B. WINSTEN (with Discussion)—Behaviour Sequences as Semi-Markov Chains. By VIOLET R. CANE (with Discussion)—On a Multivariate Version of Fieller's Theorem. By B. M. BENNETT—The Process Curve and the Equivalent Mixed Binomial with Two Components. By M. K. VAGHOLKAR—The Information in an Experiment. By C. L. MALLOWS—The Behrens-Fisher Distribution and Weighted Means. By G. S. JAMES—The Estimation of Relationships with Autocorrelated Residuals by the Use of Instrumental Variables. By J. D. SARGAN—On Some Problems of Machine Interference. By S. K. NASR—On the Distribution of Various Sums of Squares in an Analysis of Variance Table for Different Classifications with Correlated and Non-Homogeneous Errors. By B. R. BHAT—Contagious Occupancy. By D. E. BARTON and F. N. DAVID—The Limiting Frequencies of Integers with a Given Partitional Characteristic. By W. F. BODMER—Some Simple Duration-Dependent Stochastic Processes. By A. MERCER—Cyclic Queues with Feedback. By P. D. FINCH—A Statistical Theory of Remnants. By J. AITCHISON—Bandwidth and Resolvability in Statistical Spectral Analysis. By Z. A. LOMNICKI and S. K. ZAREMBA—On a Property of Incomplete Blocks. By R. MORLEY JONES—A Renewal Problem with Bulk Ordering of Components. By D. R. COX—The Dispersion of a Number of Species. By D. E. BARTON and F. N. DAVID—On a Discriminatory Problem Connected with the Works of Plato. By D. R. COX and L. BRANDWOOD—Experiments with Mixtures. By M. H. QUENOUILLE—A Different Loss Function for the Choice between Two Populations. By RITA J. MAURICE—Censored Observations in Randomized Block Experiments. By M. R. SAMPFORD and J. TAYLOR—Corrigenda: Dilution Series: A Statistical Test of Technique. By W. L. STEVENS—The Regression Analysis of Binary Sequences. By D. R. COX.

ROYAL STATISTICAL SOCIETY, 21 BENTINCK STREET, LONDON, W.1

APPLIED STATISTICS

A JOURNAL OF THE ROYAL STATISTICAL SOCIETY

VOLUME VII, No. 3

CONTENTS

NOVEMBER 1958

Analysis of Variance of a Randomised Block Design with Missing Observations. WILLIAM A. GLENN and CLYDE YOUNG KRAMER.

Professor Hogben's 'Crisis'—A Survey of the Foundations of Statistics. D. V. LINDLEY.

A Guide to Acceptance Sampling. LEONARD H. C. TIPPETT.

Some Basic Principles of Sampling Inspection by Attributes. HUGO C. HAMAKER.

The Lognormal Distribution in Quality Control. JAMES MORRISON.

NOTES AND COMMENTS.

MEETINGS OF SECTIONS OF THE ROYAL STATISTICAL SOCIETY.

BOOK REVIEWS AND PUBLICATIONS RECEIVED.

VOLUME VIII, No. 1

MARCH 1959

Some Aspects of Quality Control in Cotton Textiles. GEOFFREY NEWBERY.

The Use of the Same Experimental Material for more than One set of Treatments. G. H. FREEMAN.

Elimination of Bias in the Preliminary Examination of Data. DENIS H. WARD.

The Pattern of Consumer Purchases. A. S. C. EHRENBERG.

Choice of Conversion Factor in the Derivation of an Index of a Discontinued Series. K. S. BANERJEE.

A Significance Chart for Percentages. SIDNEY ROSENBAUM.

NOTES AND COMMENTS.

MEETINGS OF SECTIONS OF THE ROYAL STATISTICAL SOCIETY.

BOOK REVIEWS AND PUBLICATIONS RECEIVED.

APPLIED STATISTICS is published three times per year. The annual subscription is £1. 10s. (U.S.A. and Canada \$5.00). Single copies 13s. post free. Orders should be sent to

Oliver and Boyd Ltd., Tweeddale Court, 14 High Street, Edinburgh, 1

ON THE CUMULANTS OF RENEWAL PROCESSES

BY WALTER L. SMITH†

DEDICATED TO THE MEMORY OF JOHN WISHART

1. INTRODUCTION

Let $X_1, X_2, X_3, \dots$ be a *renewal process*; that is, a sequence of independent, non-negative, and identically distributed random variables which are not all zero with probability one. If there is a real, strictly positive $\tilde{\omega}$ such that, with probability one, all the X_i are divisible by $\tilde{\omega}$, then we shall call the renewal process *discrete*. A renewal process which is not discrete is termed *continuous*. Write $F(\cdot)$ for the distribution function of X_1 and $\mu_r = \mathcal{E}X_1^r (\leq \infty)$, $S_k = X_1 + X_2 + \dots + X_k$ and $F_k(\cdot)$ for the distribution function of S_k .

Renewal theory is concerned primarily with the random variable N_t which is defined to be the maximum suffix k for which $S_k \leq t$, subject to the convention $N_t = 0$ if $X_1 > t$. In the obvious renewal application the X_i represent successive lifetimes of the object being renewed, and N_t is the number of renewals made by time t (subject to the original object having been installed at time 0). However, renewal theory has proved to be of significance in a wide range of investigations. We mention the theory of queues (Smith, 1953; Beneš, 1957), the theory of electronic counters (Takács, 1956; Smith, 1957), the theory of regenerative stochastic processes (Smith, 1955), the study of the variance-time curve, in its application to problems of nerve physiology (Cox & Smith, 1954; Cox, 1955), and the theory of cricket scores (Skellam & Shenton, 1957).

Previous theoretical work on renewal theory has been concerned almost exclusively (for very good reasons) with the first moment $\mathcal{E}N_t$. We refer to papers by Feller (1941), (1949), Täcklind (1945), Blackwell (1948), and Smith (1954). A typical result concerning $\mathcal{E}N_t$ is that if the renewal process is continuous and $\mu_2 < \infty$, then as $t \rightarrow +\infty$,

$$\mathcal{E}N_t = \frac{t}{\mu_1} + \left(\frac{\mu_2}{2\mu_1^2} - 1 \right) + o(1).$$

This theorem was first proved, under slightly unnecessary restrictions, by Täcklind (1945), and in its final form by Smith (1954) (see also, Smith, 1957).

The study of higher moments of N_t seems to have been neglected, apart from a rather non-rigorous and short account by Skellam & Shenton (1957). Feller (1949) proves a theorem concerning the variance of N_t in the discrete renewal process, and Smith (1954) proves the corresponding result for continuous renewal processes: if $\mu_2 < \infty$ then as $t \rightarrow +\infty$

$$\text{var } N_t = \frac{\mu_2 - \mu_1^2}{\mu_1^3} t + o(t).$$

A result of Cox & Smith (1954) concerning a related 'equilibrium' renewal process also suggests that if $\mu_3 < \infty$ then there are constants A and B such that as $t \rightarrow +\infty$, $\text{var } N_t = At + B + o(1)$. A proof of such a result does not seem to be available, and nothing appears to have been established concerning the higher cumulants of N_t . It is with these higher cumulants that the present paper is concerned.

† This paper was prepared with the partial support of the Office of Naval Research.

The theorems concerning $\mathcal{E}N_t$ which are mentioned above have all been proved without irrelevant assumptions concerning $F(\cdot)$. However, reference to Smith (1954) will show that a very special type of argument was needed to prove that $\text{var } N_t \sim At$, which shows no prospect of generalizing to higher moments of N_t , nor of yielding the linear approximation $At + B$. Write $\mathcal{S}$ for the class of distribution functions $F(\cdot)$ for which $F_k(\cdot)$, for some finite k , has an absolutely continuous component. The present discussion of the higher cumulants of N_t , for continuous renewal processes, will be based on the weak, but extraneous, assumption that $F(\cdot) \in \mathcal{S}$. *This assumption is assumed to hold throughout this paper even when not mentioned.* It will be noticed that once this extra assumption has been made the investigation proceeds with little real difficulty. Moreover, the fact that $F(\cdot) \in \mathcal{S}$ will be found to imply certain variational properties of remainder terms that arise.

The paper falls naturally into two parts. The first part (§§ 2–5) is concerned with the rigorous proof that the n th cumulant of N_t has a linear asymptotic form, under general conditions. Two theorems (Theorems 3 and 4) are proved, which are jointly equivalent to the following.

THEOREM 0. *If $F(\cdot) \in \mathcal{S}$ and $\mu_{n+p+1} < \infty$, $p \geq 0$, then there exist constants a_n and b_n such that the n th cumulant of N_t is given by*

$$a_n t + b_n + \frac{\lambda(t)}{(1+t)^p},$$

where $\lambda(t)$ is of bounded total variation, is $o(1)$ as $t \rightarrow +\infty$, satisfies the condition

$$\lambda(t) - \lambda(t - \alpha) = o(t^{-1})$$

as $t \rightarrow +\infty$ for every fixed $\alpha > 0$, and when $p \geq 1$ has the additional property that $\lambda(t)/(1+t)$ belongs to the class L_1 .

This theorem substantially generalizes Theorems 8, 9, and 10 of Smith (1954) which were concerned with different aspects of Theorem 0 in the case $n = 1$.

The second part of the paper (§§ 6, 7) is concerned with the evaluation of a_n and b_n in terms of the μ_r . Their values are given for $n = 1, 2, \dots, 8$.

In § 2 we introduce certain unconventional moments and cumulants which are more convenient in the present investigation than the conventional ones, and we study their properties. In § 3 a number of preliminary lemmas are proved, including in particular Lemma 4 which implies a property of the remainder term in Taylor expansions of characteristic functions which we believe to be of independent interest. In § 4 we develop a symbolic calculus for dealing with complicated Laplace transforms that arise and use this calculus to prove Theorems 1 and 2 concerning the asymptotic behaviour of moments of N_t . These results are preparatory to § 5 in which Theorem 0, above, is essentially proved.

In § 6 we show how the a_n and b_n , mentioned above, may be computed. The method we used is described in sufficient detail to make it possible to extend the list of formulae given for $\{a_n\}$ and $\{b_n\}$, if required. In § 7 various checking systems are discussed which assure us that the calculations of § 6 are correct. A special case which we will quote here, and which is a consequence of Theorem 0 and the computations in §§ 6 and 7, is as follows: If $\mu_3 < \infty$ then as $t \rightarrow +\infty$,

$$\text{var } N_t = \frac{\mu_2 - \mu_1^2}{\mu_1^3} t + \left(\frac{5\mu_2^2}{4\mu_1^4} - \frac{2\mu_3}{3\mu_1^3} - \frac{\mu_2}{\mu_1^2} \right) + o(1).$$

The asymptotic linearity of the cumulants of N_t implies a curious combinatorial identity, which we have been unable to prove directly. We discuss this identity in an appendix. If it

could be established by a direct and elementary argument, then part of the proof of Theorem 0 could be drastically simplified.

Notice that this entire investigation is exclusively concerned with continuous renewal processes. Nevertheless, there is a parallel treatment for the discrete process, in which the requirement $F(\cdot) \in \mathcal{S}$ can (and, of course, must) be dropped.

The author is grateful to his friend Dennis V. Lindley, who asked the questions which provoked the present work. He also wishes to thank Professor J. Neyman for making it possible for him to spend the summer of 1957 at Berkeley, where the bulk of this paper was prepared.

2. SOME MOMENT FORMULAE

Define the r th conventional moment $m_r(t) = \mathcal{E}N_t^r$ and write $k_r(t)$ for the corresponding conventional cumulant of order r . These conventional moments and cumulants are not the most suitable for a discussion of N_t . Instead, we shall use the moments

$$\phi_k(t) = \mathcal{E}(N_t + 1)(N_t + 2) \dots (N_t + k) \quad (2.1)$$

which come from the generating function

$$\Phi_t(\zeta) = \mathcal{E} \frac{1}{(1 - \zeta)^{N_t + 1}}.$$

The cumulants $\psi_k(t)$ are obtained by a formal power series expansion of the cumulant generating function $\Psi_t(\zeta) = \log \Phi_t(\zeta)$ in the usual way. Thus, the relations between the ϕ -moments and the ψ -cumulants are exactly similar to the corresponding relations between the conventional moments $m_r(t)$ and the conventional cumulants $k_r(t)$. It is evident, from (2.1), that

$$\begin{aligned} \phi_1(t) &= m_1(t) + 1, \\ \phi_2(t) &= m_2(t) + 3m_1(t) + 2, \\ \phi_3(t) &= m_3(t) + 6m_2(t) + 11m_1(t) + 6, \\ \phi_4(t) &= m_4(t) + 10m_3(t) + 35m_2(t) + 50m_1(t) + 24. \end{aligned}$$

In fact, it can easily be shown that

$$\phi_n(t) = |S_{n+1}^{n+1}| m_n(t) + |S_{n+1}^n| m_{n-1}(t) + |S_{n+1}^{n-1}| m_{n-2}(t) + \dots + |S_{n+1}^1|, \quad (2.2)$$

where S_n^m are Stirling's numbers of the first kind (Jordan, 1947, p. 142).

By systematic solution of equations (2.2) one can deduce the reciprocal relations

$$\begin{aligned} m_1(t) &= \phi_1(t) - 1, \\ m_2(t) &= \phi_2(t) - 3\phi_1(t) + 1, \\ m_3(t) &= \phi_3(t) - 6\phi_2(t) + 7\phi_1(t) - 1, \\ m_4(t) &= \phi_4(t) - 10\phi_3(t) + 25\phi_2(t) - 15\phi_1(t) + 1. \end{aligned}$$

However, it is not difficult to deduce from the defining property of the Stirling numbers, $\mathfrak{S}_n^m$, of the second kind (Jordan, 1947, p. 168) that the general relation is

$$m_n(t) = \mathfrak{S}_{n+1}^{n+1} \phi_n(t) - \mathfrak{S}_{n+1}^n \phi_{n-1}(t) + \dots + (-1)^n \mathfrak{S}_{n+1}^1. \quad (2.3)$$

Let us write $M_t(z) = \mathcal{E}e^{zN_t}$ for the conventional moment generating function, and $K_t(z) = \log M_t(z)$ for the corresponding cumulant generating function. It can be shown that $M_t(z) = e^{-z} \Phi_t(1 - e^{-z})$ and hence that $K_t(z) = -z + \Psi_t(1 - e^{-z})$. Thus we have

$$k_1(t)z + \frac{k_2(t)}{2!}z^2 + \dots = -z + \psi_1(t)[1 - e^{-z}] + \frac{\psi_2(t)}{2!}[1 - e^{-z}]^2 + \dots \quad (2.4)$$

A result of Jordan (1947, p. 202) shows that

$$\frac{(1 - e^{-z})^m}{m!} = \sum_{k=m}^{\infty} \frac{\mathfrak{S}_k^m (-1)^{m+k}}{k!} z^k; \quad (2.5)$$

and from (2.4) and (2.5) we deduce the following relations between conventional cumulants $k_n(t)$ and the ψ -cumulants $\psi_n(t)$.

$$\left. \begin{aligned} k_1(t) &= \psi_1(t) - 1, \\ k_n(t) &= \mathfrak{S}_n^n \psi_n(t) - \mathfrak{S}_{n-1}^{n-1} \psi_{n-1}(t) + \dots + (-1)^{n-1} \mathfrak{S}_n^1 \psi_1(t) \quad (n \geq 2). \end{aligned} \right\} \quad (2.6)$$

For reference we list the following:

$$\left. \begin{aligned} k_2(t) &= \psi_2(t) - \psi_1(t), \\ k_3(t) &= \psi_3(t) - 3\psi_2(t) + \psi_1(t), \\ k_4(t) &= \psi_4(t) - 6\psi_3(t) + 7\psi_2(t) - \psi_1(t). \end{aligned} \right\} \quad (2.7)$$

Finally, to complete our discussion of the relations between the $k_n(t)$ and the $\psi_n(t)$, let us notice that

$$\Psi_l(\zeta) = \log(1 - \zeta) + K_l(-\log[1 - \zeta]). \quad (2.8)$$

If we use the following formula of Jordan ((1947), p. 202),

$$\frac{[\log(1 + \zeta)]^m}{m!} = \sum_{n=m}^{\infty} \frac{S_n^m}{n!} \zeta^n,$$

then we can deduce from (2.8) that

$$\psi_r(t) = |S_r^r| k_r(t) + |S_r^{r-1}| k_{r-1}(t) + \dots + |S_r^2| k_2(t) + |S_r^1| [k_1(t) - 1].$$

Whilst this last formula is occasionally useful for computational checks it will have no explicit use in the present paper.

3. SOME NOTATION AND PRELIMINARY LEMMAS

All functions with which we shall be concerned will be assumed to vanish identically for negative values of their arguments. If $\Omega(t)$ is any function of bounded variation in every finite interval then we shall write

$$*\Omega(s) = \int_{0-}^{\infty} e^{-st} d\Omega(t)$$

for the Laplace-Stieltjes transform (L.-S.T.) of $\Omega(t)$. If $\Omega(t)$ is any function which is integrable in every finite interval then we write

$$*\Omega(s) = \int_0^{\infty} e^{-st} \Omega(t) dt$$

for the Laplace transform (L.T.) of $\Omega(t)$. The transforms that arise in the present work will all exist for $\Re s > 0$. Notice that when $\Omega(t)$ has both a L.-S.T. and a L.T. then $*\Omega(s) = s_*\Omega(s)$.

The following lemmas concerning $*F(s)$ undoubtedly include much that is known. However, there does not seem to be available a satisfactory and complete treatment, dealing specifically with the one-sided Laplace transform, which leads to the somewhat surprising Lemma 4, below. For this reason we give the following proofs in some detail.

LEMMA 1. If $\mu_n < \infty$ and

$$\theta_n^s(x) = e^{-sx} \int_0^x e^{sz} z^n dz$$

then, as $s \rightarrow 0+$,

$$\int_{0-}^{\infty} \theta_n^s(x) dF(x) = o(s^{-1}).$$

Proof. For any $\epsilon > 0$ we can find an $x_0(\epsilon)$ such that

$$\int_{x_0}^{\infty} z^n dF(z) < \epsilon,$$

since $\mu_n < \infty$. But

$$\begin{aligned} \theta_n^s(x) &< x^n \int_0^x e^{-sz} dz \\ &< x^n/s. \end{aligned}$$

Hence,

$$\int_{x_0}^{\infty} \theta_n^s(z) dF(z) < \epsilon/s. \quad (3.1)$$

Also, for $x \leq x_0$,

$$\theta_n^s(x) \leq \frac{x_0^n [1 - e^{-sx_0}]}{s},$$

and for $s < s_0(x_0(\epsilon), \epsilon)$,

$$x_0^n [1 - e^{-sx_0}] < \epsilon.$$

Thus

$$\int_{0-}^{x_0} \theta_n^s(z) dF(z) < \epsilon/s. \quad (3.2)$$

The lemma follows from (3.1) and (3.2).

LEMMA 2.

$$e^{-\alpha x} = 1 - \alpha x + \frac{\alpha^2 x^2}{2!} - \dots + \frac{(-\alpha x)^n}{n!} + \frac{(-\alpha)^{n+1}}{n!} \theta_n^{\alpha}(x). \quad (3.3)$$

Proof. The L.T. of the right-hand side of (3.3) is

$$\frac{1}{s} - \frac{\alpha}{s^2} + \frac{\alpha^2}{s^3} - \dots + \frac{(-\alpha)^n}{s^{n+1}} + \frac{(-\alpha)^{n+1}}{(s+\alpha)s^{n+1}} = \frac{1}{s+\alpha}.$$

LEMMA 3. A necessary and sufficient condition for $F(x)$ to have its first n moments finite is that, for real $s > 0$, $*F(s)$ have an expansion

$$*F(s) = 1 - \mu_1 s + \frac{\mu_2 s^2}{2!} - \dots + \frac{(-s)^n \mu_n}{n!} + o(s^n).$$

Proof. Suppose first that $\mu_1, \mu_2, \dots, \mu_n$ are finite. Then, by Lemmas 1 and 2 we have

$$\begin{aligned} *F(s) &= \int_{0-}^{\infty} e^{-sz} dF(z) \\ &= \int_{0-}^{\infty} \left\{ 1 - sz + \frac{s^2 z^2}{2!} - \dots + \frac{(-s)^n}{n!} z^n + \frac{(-s)^{n+1}}{n!} \theta_n^s(z) \right\} dF(z) \\ &= 1 - \mu_1 s + \frac{\mu_2}{2!} s^2 - \dots + \frac{\mu_n}{n!} (-s)^n + s^{n+1} o\left(\frac{1}{s}\right). \end{aligned}$$

This proves the necessity half of the theorem.

To prove the sufficiency part define the shift operator E^h by the rule $E^h g(s) = g(s+h)$, for any function $g(s)$. Then the k th derivative of $*F(s)$ on the right is given by

$$*F^{(k+)}(s) = \lim_{h \rightarrow 0+} \left(\frac{E^h - 1}{h} \right)^k *F(s).$$

If the expansion of $*F(s)$ announced in the statement of our lemma holds, then for $k \leq n$ we have

$$*F^{(k+)}(0) = \mu_k < \infty.$$

However,

$$\begin{aligned} *F^{(k+)}(s) &= \text{Lt}_{h=0+} \left(\frac{E^h - 1}{h} \right)^k \int_{0-}^{\infty} e^{-sz} dF(z), \\ &= \text{Lt}_{h=0+} \int_{0-}^{\infty} \left(\frac{e^{-hz} - 1}{h} \right)^k e^{-sz} dF(z). \end{aligned}$$

Thus

$$\begin{aligned} \mu_k &= \text{Lt}_{h=0+} \int_{0-}^{\infty} \left(\frac{e^{-hz} - 1}{hz} \right)^k z^k dF(z) \\ &\geq \lim_{h=0+} \frac{1}{2^k} \int_{0-}^{\epsilon h^{-1}} z^k dF(z), \end{aligned}$$

where ϵ is chosen so that for all $0 \leq z \leq \epsilon$ we have $1 - e^{-z} \geq \frac{1}{2}z$. The finiteness of μ_k (regarded, temporarily, purely as a coefficient in the given expansion of $*F(s)$) establishes the finiteness of

$$\int_{0-}^{\infty} z^k dF(z).$$

LEMMA 4. *A necessary and sufficient condition for the distribution function $F(x)$ of a non-negative random variable to have its first k moments finite is that there exists another distribution function $F_{(1)}(x)$ such that*

$$*F(s) = 1 - \mu_1 s + \frac{\mu_2}{2!} s^2 - \dots + \frac{\mu_{k-1}}{(k-1)!} (-s)^{k-1} + \frac{\mu_k}{k!} (-s)^k + *F_{(1)}(s).$$

Proof. Obviously, if the above expansion is possible, then Lemma 3 and the fact that $*F_{(1)}(s) = 1 + o(1)$ as $s \rightarrow 0+$ prove the finiteness of the first k moments of $F(x)$.

Now suppose that we are merely given the fact that $\mu_k < \infty$. We shall assume $k \geq 1$ since the lemma is trivial if $k = 0$. If the mean μ_1 is finite then we can define a new distribution function

$$F_{(1)}(x) = \frac{1}{\mu_1} \int_0^x \{1 - F(z)\} dz,$$

with L.-S.T.

$$*F_{(1)}(s) = \frac{1 - *F(s)}{\mu_1 s}.$$

From Lemma 3 we see that

$$*F(s) = 1 - \mu_1 s + \frac{\mu_2}{2!} s^2 - \dots + \frac{\mu_k}{k!} (-s)^k + o(s^k),$$

so that

$$*F_{(1)}(s) = 1 - \frac{\mu_2}{2! \mu_1} s + \frac{\mu_3}{3! \mu_1} s^2 - \dots + \frac{\mu_k}{k! \mu_1} (-s)^{k-1} + o(s^{k-1}).$$

It then follows, also from Lemma 3, that the first $(k-1)$ moments of $F_{(1)}(x)$ are finite, and that the r th of these is given by $\mu_{r+1}/(r+1)\mu_1$. Notice, moreover, that

$$*F(s) = 1 - \mu_1 s *F_{(1)}(s).$$

We can now repeat the process, e.g. define

$$F_{(2)}(x) = \frac{2\mu_1}{\mu_2} \int_0^x \{1 - F_{(1)}(z)\} dz,$$

and obtain a new distribution function $F_{(2)}(x)$ whose first $(k-2)$ moments are finite, and given by the expansion

$$*F_{(2)}(s) = 1 - \frac{2\mu_3}{3!\mu_2}s + \frac{2\mu_4}{4!\mu_2}s^2 - \dots + \frac{2\mu_k}{k!\mu_2}(-s)^{k-2} + o(s^{k-2}).$$

For this second 'derived' distribution function $F_{(2)}(x)$ we have

$$*F(s) = 1 - \mu_1 s + \frac{\mu_2}{2!} s^2 *F_{(2)}(s).$$

Evidently, repeated derivations of $F_{(3)}(x)$, $F_{(4)}(x)$, ..., $F_{(k)}(x)$, will prove the lemma.

Whilst strictly irrelevant to the present discussion, the remark is worth making that Lemma 4 is capable of trivial modification to include distributions of unrestricted (i.e. not necessarily positive) random variables. A final preparatory result is the following.

LEMMA 5. If $\Omega(x)$ is any function of bounded (total) variation (B.V.) then

$$*\Omega^{(k)}(s) = o(s^{-k}) \quad \text{as } s \rightarrow 0+.$$

Proof.

$$(-s)^k * \Omega^{(k)}(s) = \int_{0-}^{\infty} e^{-sx} (sx)^k d\Omega(x).$$

But, as $s \rightarrow 0+$, $e^{-sx}(sx)^k \rightarrow 0$, boundedly. This proves the lemma.

4. THE ASYMPTOTIC FORM OF $\phi_n(t)$

We need first

LEMMA 6. $*\phi_n(s) = n! \{1 - *F(s)\}^{-n}$.

Proof. Since $\phi_n(t)$ is evidently a non-decreasing function of t it is legitimate to consider the L.-S.T. $*\phi_n(s)$.

Now

$$\begin{aligned} \phi_n(t) &= \mathcal{E}(N_t + 1)(N_t + 2) \dots (N_t + n) \\ &= \sum_{k=0}^{\infty} (k+1)(k+2) \dots (k+n) \{F_k(t) - F_{k+1}(t)\}. \end{aligned}$$

Thus

$$\begin{aligned} *\phi_n(s) &= \sum_{k=0}^{\infty} (k+1)(k+2) \dots (k+n) \{[*F(s)]^k - [*F(s)]^{k+1}\} \\ &= n! \{1 - *F(s)\} \sum_{k=0}^{\infty} \frac{(k+1) \dots (k+n)}{n!} [*F(s)]^k. \end{aligned}$$

The lemma follows at once.

We remark that it is the particularly simple form of the L.-S.T. $*\phi_n(s)$ that makes $\phi_n(t)$ so much more convenient, for our present study, than the conventional moment $m_n(t)$.

Provided $\mu_n < \infty$ we can consider the function

$$*K_n(s) = \frac{1 - *F_{(n)}(s)}{1 - *F(s)} \quad (n \geq 1),$$

concerning which we have the following three lemmas

LEMMA 7. If $F(t) \in \mathcal{S}$ and $\mu_{n+1} < \infty$ then $K_n(t)$ is of bounded total variation, and

$$K_n(\infty) = \frac{\mu_{n+1}}{(n+1)\mu_1\mu_n}.$$

Proof. The fact that $K_n(t)$ is of bounded total variation follows from the fact that if $\mu_{n+1} < \infty$ then the means of both $F(t)$ and $F_{(n)}(t)$ are finite, and from Lemma 6 of Smith (1954). The mean of $F_{(n)}(t)$ is found, from the Taylor expansion of $*F_{(n)}(s)$, to be $\mu_{n+1}/(n+1)\mu_n$, and so, letting $s \rightarrow 0+$ in $*K_n(s)$ we deduce that $*K_n(0) = K_n(\infty)$ has the value announced.

LEMMA 8. If $\mu_{n+1} < \infty$, so that $*K_{n+1}(s)$ exists, then

$$\frac{*K_n(s)}{s} = \frac{\mu_{n+1}}{(n+1)\mu_n} \left[\frac{1}{\mu_1 s} + *K_1(s) - *K_{n+1}(s) \right]. \quad (4.1)$$

Proof. We know that

$$*F_{(n)}(s) = 1 - \frac{\mu_{n+1}}{(n+1)\mu_n} s *F_{(n+1)}(s).$$

But the right-hand side of (4.1) is equal to

$$\frac{\mu_{n+1}}{(n+1)\mu_n} \left[\frac{*F_{(1)}(s)}{1 - *F(s)} + \frac{1 - *F_{(1)}(s)}{1 - *F(s)} - \frac{1 - *F_{(n+1)}(s)}{1 - *F(s)} \right] = \frac{\mu_{n+1}}{(n+1)\mu_n} \frac{*F_{(n+1)}(s)}{1 - *F(s)}.$$

The lemma is thus evident.

LEMMA 9.

$$*\phi_1(s) = \frac{1}{\mu_1 s} + *K_1(s).$$

Proof. Direct computation establishes the lemma.

So far we have been able to deal exactly and precisely with the functions and coefficients which have arisen. If we were to attack in this way the theorems which we now approach we would rapidly become involved in hopeless complication. Our primary objective in this section is merely to establish that $\phi_n(t)$ has a polynomial form with a remainder term of a certain type. At this stage we shall therefore introduce, and make considerable use of, certain notations which will preserve the essence of the various relations in which we have interest, without preserving a great deal of undesired and heavy detail. We shall write γ for any constant, and γ_n for any finite rational function of $\mu_1, \mu_2, \dots, \mu_n$. For instance, we would replace $\mu_1^2 \mu_3$ by γ_3 . However, nothing is to stop us calling $\mu_1^2 \mu_3$ by the name γ_{12} , or γ_{13} , if we want to. We shall also write $*M_w^d$ for any multinomial expression in $*K_1, *K_2, \dots, *K_w$, with coefficients γ_w and with the maximum degree of any term not greater than d . We shall also insist, when $w > 1$, that the degree of $*K_w$ in any term is not to exceed unity. For examples of the use of this symbolism we offer the following.

- (i) $\mu_1^4 \mu_4 *K_3 *K_4 + \mu_2 *K_2^2 = *M_4^2,$
- (ii) $\mu_1 *K_6 *K_1^5 + \mu_6 *K_5 = *M_6^6.$

Certain points about the notation $*M_w^d$ should be noticed. The multinomial terms in $*M_w^d$ do not necessarily have to be of homogeneous degree. A $*K$ or γ with maximum allowable suffix does not necessarily have to occur. The essential feature of the notation $*M_w^d$ is that if $\mu_{w+1} < \infty$ then it represents the L.-S.T. of a function of B.V. This is because any γ_w is certainly defined and finite if $\mu_{w+1} < \infty$, every $*K_r(s)$ with $r \leq w$ will be (by Lemma 7) the L.-S.T. of a function of B.V., and the product of a finite number of L.-S.T.'s of functions of B.V. is itself the L.-S.T. of a function of B.V. Incidentally, it is easy to see from Lemma 7 that $M_w^d(+\infty)$, which is the same thing as $*M_w^d(0)$, the total net variation of $M_w^d(t)$, is simply γ_{w+1} .

Expressions such as $\gamma_3, *M_4^2$, etc., may occur several times in the same equation; but it must not be supposed that at each appearance they represent the same function. However,

the following laws of combination should be noticed. If $r = \max(p, q)$ and $d = \max(b, c)$, then

$$\gamma_p + \gamma_q = \gamma_r, \quad (C1)$$

$$\gamma_p \gamma_q = \gamma_r, \quad (C2)$$

$$*M_p^b + *M_q^c = *M_r^d, \quad (C3)$$

$$\left. \begin{aligned} *M_p^b *M_q^c &= *M_{r+1}^{b+c} \quad \text{if } p = q \neq 1 \\ &= *M_r^{b+c} \quad \text{otherwise.} \end{aligned} \right\}, \quad (C4)$$

The slight complication in rule (C4) arises from the fact that in a multinomial $*M_w^d$ with $w > 1$, a $*K_w$ may not have a higher degree than unity in any term. When we multiply two multinomials, of the same weight w , it is possible for a term containing a factor $*K_w^2$ to arise.

It should be clear that we may consistently write

$$*M_w^0 = \gamma_w.$$

LEMMA 10. If $d \geq 1$, then

$$\frac{*M_w^d}{s} = \frac{\gamma_{w+1}}{s} + *M_{w+1}^d.$$

Proof. A typical term in $*M_w^d$ could be written as $\gamma_w *K_{p_1} *K_{p_2} \dots *K_{p_r}$, where $r \leq d$ and every $p_j \leq w$. This term could be written, alternatively, as $*K_{p_1} *M_w^{d-1}$. But Lemma 8 states, essentially, that

$$\frac{*K_{p_1}}{s} = \frac{\gamma_{p_1+1}}{s} + *M_{p_1+1}^1.$$

Thus, remembering that $p_1 \leq w$, we have by (C4) that

$$\frac{*K_{p_1} *M_w^{d-1}}{s} = \frac{\gamma_{w+1} *M_w^{d-1}}{s} + *M_{w+1}^d.$$

From this symbolic reduction of a typical term in $*M_w^d$ it is thus apparent that

$$\frac{*M_w^d}{s} = \gamma_{w+1} \frac{*M_w^{d-1}}{s} + *M_{w+1}^d. \quad (4.2)$$

If $d = 1$ then the lemma is proved, since $*M_w^0 = \gamma_w$. If not, we re-apply the above argument to $*M_w^{d-1}$, and deduce

$$\frac{*M_w^{d-1}}{s} = \gamma_{w+1} \frac{*M_w^{d-2}}{s} + *M_{w+1}^{d-1}. \quad (4.3)$$

On combining (4.2) and (4.3), and using (C2) and (C3), we have

$$\frac{*M_w^d}{s} = \gamma_{w+1} \frac{*M_w^{d-2}}{s} + *M_{w+1}^{d-1} + *M_{w+1}^d = \gamma_{w+1} \frac{*M_w^{d-2}}{s} + *M_{w+1}^d. \quad (4.4)$$

If $d = 2$ the lemma is proved at this stage, and it should now be clear that repeated reductions will eventually prove the lemma, whatever the value of d .

In our discussions of remainder terms we shall find it convenient to use the following classes of functions, which we now define. All functions, as has been stated before, are assumed to vanish for negative argument.

DEFINITION 1. The function $\lambda(t)$ belongs to the class $\mathcal{B}$ if and only if it is B.V., tends to zero as t approaches $+\infty$, and satisfies the condition

$$\lambda(t) - \lambda(t - \alpha) = o(t^{-1}), \quad (*)$$

as $t \rightarrow +\infty$, for every $\alpha > 0$.

DEFINITION 2. The function $\lambda(t)$ belongs to the class $\mathcal{L}$ if and only if $\lambda(t)/(1+t)$ belongs to L_1 .

It will be convenient to write $\mathcal{R} = \mathcal{B} \cap \mathcal{L}$ for the class of all those functions in both $\mathcal{B}$ and $\mathcal{L}$. An easy result which we use later is

$$\text{If } \lambda_1(t) \in \mathcal{B} \text{ and } \lambda_2(t) \in \mathcal{R}, \text{ then } \lambda_1(t) \lambda_2(t) \in \mathcal{R}. \quad (C5)$$

(In fact $\mathcal{B}$ is a commutative algebra over the reals and (C5) is included in the remark that $\mathcal{R}$ is an ideal in $\mathcal{B}$. However, these concepts will find no application in the present work.)

LEMMA 11. If $\mu_w < \infty$, $w > 1$, then

(i) $M_w^d(t) = o(t)$ as $t \rightarrow \infty$,

(ii) whenever $G(t)$ is B.V., the function $P(t)$ whose L.-S.T. is $*G * M_w^d$ satisfies the conditions

$$P(t) - P(t - \alpha) = o(1) \text{ as } t \rightarrow \infty, \text{ for every } \alpha > 0. \quad (**)$$

Proof. Terms in $*M_w^d$ may contain $*K_w$ to at most the first power so that we can evidently write $*M_w^d$ as $*K_w *J + *L$, where $*J$ and $*L$ contain only $*K_p$ terms with $p < w$. Since $\mu_w < \infty$, it follows that $J(t)$ and $L(t)$ are both functions of B.V. Moreover, if we write $*\Psi = *J\{1 - *F_w\}$ then $\Psi(t)$ is also a function of B.V. and is, indeed, one which vanishes at $+\infty$ (since $*\Psi(0+) = 0$ is obviously zero). Rewriting $*M_w^d$ as $*L + *\Psi * \phi_1$ we see that

$$M_w^d(t) = L(t) + \int_0^t \Psi(t-t') d\phi_1(t'),$$

and the fact that $M_w^d(t) = o(t)$ now follows from Lemma 1 of Smith (1955).

To prove the second part of the lemma we notice that the L.-S.T. of $P(t) - P(t - \alpha)$ is simply

$$\{1 - e^{-\alpha s}\} *G * M_w^d = \{1 - e^{-\alpha s}\} *G * L + \frac{1 - e^{-\alpha s}}{1 - *F} *G * \Psi,$$

which we recognize, by Lemma 6 of Smith (1954), to be the L.-S.T. of a function of B.V. Evidently this function approaches zero as $t \rightarrow +\infty$, as may be seen by letting $s \rightarrow 0+$. Thus the lemma is proved.

LEMMA 12. For any non-negative integer p , define $*g_p(s) = s^{-p} *M_w^d$. Then, if $\mu_{w+1} < \infty$, $w \geq 1$, we can find a γ_{w+1} such that $t^{-p}g_p(t) - \gamma_{w+1}$ belongs to $\mathcal{B}$.

Proof. Evidently

$$g_p(t) = \int_0^t (t-z)^p dM_w^d(z),$$

where M_w^d is B.V. Hence

$$\frac{g_p(t)}{t^p} = M_w^d(t) + \sum_{r=1}^{p-1} \int_0^t \left(\frac{z}{t}\right)^r dM_w^d(z).$$

This suggests that we take for the γ_{w+1} referred to in the enunciation the value $M_w^d(+\infty)$. Since $M_w^d(t)$ is B.V. we shall have proved that $t^{-p}g_p(t) - M_w^d(+\infty)$ is B.V., and $o(1)$ at $+\infty$, if we can prove that

$$l_r(t) = \frac{1}{t^r} \int_0^t z^r dV(z)$$

is B.V. and vanishes at $+\infty$, for every integer $r > 0$, and every distribution function $V(z)$.

Choose $\epsilon > 0$, and take T_1 so that $V(T_1) > 1 - \epsilon$. Then for all $t > T_2 > \epsilon^{-1} T_1$ it is clear that

$$\begin{aligned} l_r(t) &< \frac{1}{T_2} \int_0^{T_1} z^r dV(z) + \int_{T_1}^t dV(z) \\ &< \epsilon^r + \epsilon. \end{aligned}$$

Thus $l_r(t)$ does in fact vanish at $+\infty$. Moreover, an integration by parts will quickly verify that

$$l_r(t) = V(t) - r \int_0^t \frac{l_r(z)}{z} dz.$$

But, because $l_r(t)$ is non-negative, this displays $l_r(t)$ to be the difference between two increasing functions which must, since $V(t) \leq 1$, be bounded. Hence $l_r(t)$ is B.V.

To finish the proof we must show that the condition (*) used in Definition 1 is satisfied by $t^{-p}g_p(t)$. To this end we remark that an easy deduction from Lemma 10 is that

$$\frac{{}^*M_w^d}{s^p} = \frac{M_w^d(+\infty)}{s^p} + \frac{{}^*M_{w+1}^d}{s^{p-1}}.$$

Hence (4.5)

$$\frac{g_p(t)}{t^p} - M_w^d(+\infty) = \frac{1}{t} \int_0^t \left(1 - \frac{z}{t}\right)^{p-1} dM_{w+1}^d(z),$$

where $M_{w+1}^d(z)$ cannot be assumed to be B.V. However, let us define

$$m(t) = \int_0^t \left(1 - \frac{z}{t}\right)^{p-1} dM_{w+1}^d(z). \quad (4.6)$$

We claim that the proof will be complete if we prove that $m(t)$ satisfies the condition

$$m(t) - m(t-\alpha) = o(1) \quad \text{as } t \rightarrow +\infty, \quad \text{for every } \alpha > 0. \quad (**)$$

For, by (4.5) and the first part of this proof, we have already shown that $m(t) = o(t)$ as $t \rightarrow +\infty$. Thus $\{1 - t/(t-\alpha)\} m(t-\alpha) = o(1)$, and so, if (**) holds then (*) is satisfied by $m(t)/t$ (which is the result required).

To reduce our problem yet further we can now see from (4.6) that all we need prove is that if, for any $r \geq 0$, we define

$$n_r(t) = \frac{1}{t^r} \int_0^t z^r dM_{w+1}^d(z),$$

then $n_r(t)$ satisfies (**).

An integration by parts (identical with one performed already in this proof) shows that

$$\frac{n_r(t)}{t} + \frac{r}{t} \int_0^t \frac{n_r(z)}{z} dz = \frac{M_{w+1}^d(t)}{t}. \quad (4.7)$$

Lemma 11 (i) has proved that the right-hand side of (4.7) tends to zero as $t \rightarrow +\infty$, and so we can deduce from the integral version of Mercer's theorem (Paley & Wiener (1934) p. 58) that $n_r(t) = o(t)$ as $t \rightarrow +\infty$. Hence

$$\lim_{t \rightarrow +\infty} \int_{t-\alpha}^t \frac{n_r(z)}{z} dz = 0. \quad (4.8)$$

Multiply (4.7) by t . Then the fact that $n_r(t)$ satisfies (**) follows from the resulting equation, (4.8), and the result that $M_{w+1}^d(t)$ satisfies (**), which was proved in Lemma 11 (ii).

THEOREM 1. If $\mu_{n+1} < \infty$ then

$$\phi_n(t) = \gamma_1 t^n + \gamma_2 t^{n-1} + \dots + \gamma_n t + \gamma_{n+1} + \omega(t),$$

where $\omega(t) \in \mathcal{B}$.

Proof. By Lemma 6

$$\begin{aligned} {}^*\phi_n(s) &= n! [{}^*\phi_1(s)]^n \\ &= n! \left[\frac{\gamma_1}{s} + {}^*K_1 \right]^n, \quad \text{by Lemma 9,} \\ &= \frac{\gamma_1}{s^n} + \frac{{}^*M_1^1}{s^{n-1}} + \frac{{}^*M_1^2}{s^{n-2}} + \dots + {}^*M_1^n. \end{aligned}$$

(It is to be recalled that the rule excluding powers of ${}^*K_w$ greater than unity in ${}^*M_w^d$ does not apply when $w = 1$.)

Let us now apply Lemma 10 repeatedly. We deduce that

$$\begin{aligned} {}^*\phi_n(s) &= \frac{\gamma_1}{s^n} + \frac{\gamma_2}{s^{n-1}} + \frac{{}^*M_2^2}{s^{n-2}} + \dots + {}^*M_2^n, \\ &= \frac{\gamma_1}{s^n} + \frac{\gamma_2}{s^{n-1}} + \frac{\gamma_3}{s^{n-2}} + \frac{{}^*M_3^3}{s^{n-3}} + \dots + {}^*M_3^n, \\ &= \text{etc.}, \\ &= \frac{\gamma_1}{s^n} + \frac{\gamma_2}{s^{n-1}} + \dots + \frac{\gamma_n}{s} + {}^*M_n^n. \end{aligned} \tag{4.9}$$

Thus it follows that

$$\phi_n(t) = \gamma_1 t^n + \gamma_2 t^{n-1} + \dots + \gamma_n t + M_n^n(t).$$

The property claimed for the remainder $\omega(t)$ follows immediately from Lemma 12, in which we let $p = 0$.

LEMMA 13. If $\mu_{n+p} < \infty$, $n > 0$, $p > 0$, then

$$(-1)^p \frac{{}^*F^{(p)}(s)}{\mu_p}$$

is the L.-S.T. of a distribution function $F^{[p]}(x)$, say, whose first n moments are finite.

Proof. Evidently

$$(-1)^p {}^*F^{(p)}(s) = \int_{0-}^{\infty} e^{-sx} x^p dF(x),$$

and so, if we put

$$dF^{[p]}(x) = \frac{x^p dF(x)}{\mu_p}$$

the lemma is obvious.

THEOREM 2. If $\mu_{n+p+1} < \infty$, $n > 0$, $p > 0$, then

$$\phi_n(t) = \gamma_1 t^n + \gamma_2 t^{n-1} + \dots + \gamma_{n+1} + \frac{\lambda(t)}{(1+t)^p},$$

where $\lambda(t) \in \mathcal{R}$.

Remark. This theorem bears a close superficial resemblance to the previous theorem, but provides us with much more information about the asymptotic behaviour of the remainder term, and requires a somewhat more intricate proof. We need the extra information concerning the remainder term if we are to discuss the asymptotic behaviour of the cumulants, which involve *products* of moments.

Proof. The proof is based on the fact that $(-1)^p {}^*\phi_n^{(p)}(s)$ is the L.T. of $t^p \phi_n(t)$. However, before we enter the main part of this proof we shall establish the preliminary result

$$(-1)^p {}^*\phi_n^{(p)}(s) = \frac{\gamma_1}{s^{n+p+1}} + \frac{\gamma_2}{s^{n+p}} + \dots + \frac{\gamma_{n+1}}{s^{p+1}} + o\left(\frac{1}{s}\right). \quad (4.10)$$

To prove this preliminary result, we observe that Lemma 10 leads easily to

$$\frac{{}^*M_n^n}{s} = \frac{\gamma_{n+1}}{s} + \gamma_{n+2} + \gamma_{n+3}s + \dots + \gamma_{n+p}s^{p-2} + s^{p-1} {}^*M_{n+p}^n. \quad (4.11)$$

However, Lemma 5 and some repeated differentiation will demonstrate that

$$\left(-\frac{d}{ds}\right)^p s^{p-1} {}^*M_{n+p}^n = o\left(\frac{1}{s}\right),$$

so that we can deduce from (4.11) the fact that

$$\left(-\frac{d}{ds}\right)^p \frac{{}^*M_n^n}{s} = \frac{\gamma_{n+1}}{s^{p+1}} + o\left(\frac{1}{s}\right). \quad (4.12)$$

The preliminary result (4.10) then follows at once from (4.9) and (4.12).

Turning now to the main part of the proof we remark that by Lemma 6 we can write

$${}^*\phi_n(s) = \frac{n!}{s} \left\{ \frac{1}{1 - {}^*F(s)} \right\}^n,$$

so that $(-1)^p {}^*\phi_n^{(p)}(s)$ is the sum of a number of terms like

$$\Gamma = \gamma \frac{{}^*F^{(j_1)}(s) \dots {}^*F^{(j_r)}(s)}{s^{t+1} \{1 - {}^*F(s)\}^{n+r}}.$$

In such a term, Γ , the number of factors in the numerator is r , which is also the increase above n of the power of $\{1 - {}^*F(s)\}$ in the denominator. The multiplier γ will depend only on n , and not on the moments $\mu_1, \mu_2, \dots$ etc. Furthermore, one must have

$$j_1 + j_2 + \dots + j_r + t = p,$$

so that

$$r + t \leq p,$$

and

$$\max_i j_i \leq p - r - t + 1.$$

To substantiate these claims, consider what happens if Γ is differentiated once with respect to s . We obtain a number of terms:

- (a) one by increasing t by 1;
- (b) r terms by increasing each j_i by 1, in turn;
- (c) one term by increasing r by 1 in the denominator and adding a factor ${}^*F^{(1)}$ to the numerator.

It should now be evident that an inductive argument will prove the correctness of our assertions concerning Γ .

We shall next rewrite Γ , using Lemmas 4, 6 and 13, in the form

$$\Gamma = \gamma_p {}^*F^{[j_1]} {}^*F^{[j_2]} \dots {}^*F^{[j_r]} \{ {}^*F^{(1)} \}^t {}^*\phi_{n+r+t}.$$

This equation can be written, more simply, as

$$\Gamma = \gamma_p {}^*G(s) {}^*\phi_{n+r+t} \quad (4.13)$$

where G is a distribution function whose first

$$(n+p+1) - (p-r-t+1) = n+r+t$$

moments are finite and representable as γ_{n+p+1} . A consequence of this is that, whenever $\rho \leq n+r+t$, we can use Lemma 4 to infer that

$$\frac{{}^*G(s)}{s^{\rho+1}} = \frac{1}{s^{\rho+1}} + \frac{\gamma_{n+p+1}}{s^{\rho}} + \dots + \frac{\gamma_{n+p+1}}{s^2} + \gamma_{n+p+1} {}^*G_{[\rho]}(s).$$

If we couple this last observation with (4.9), in which n is to be replaced by $n+r+t$, then we can speedily deduce from (4.13) the representation

$$\Gamma = \frac{\gamma_{n+p+1}}{s^{n+r+t+1}} + \frac{\gamma_{n+p+1}}{s^{n+r+t}} + \dots + \frac{\gamma_{n+p+1}}{s^2} + {}^*\Omega,$$

where ${}^*\Omega$ is to represent the L.T. of a function of B.V. But Γ is a typical term in the expansion for $(-1)^p {}^*\phi_n^{(p)}(s)$, and hence we can infer that

$$(-1)^p {}^*\phi_n^{(p)}(s) = \frac{\gamma_{n+p+1}}{s^{n+p+1}} + \frac{\gamma_{n+p+1}}{s^{n+p}} + \dots + \frac{\gamma_{n+p+1}}{s^2} + {}^*\Omega, \quad (4.14)$$

where ${}^*\Omega$ continues to represent the L.T. of some (different) function of B.V.

At this stage we invoke the preliminary result (4.10) to infer that many of the coefficients in the expansion (4.14) must vanish identically, and that

$$(-1)^p {}^*\phi_n^{(p)}(s) = \frac{\gamma_1}{s^{n+p+1}} + \frac{\gamma_2}{s^{n+p}} + \dots + \frac{\gamma_{n+1}}{s^{p+1}} + {}^*\Omega.$$

Thus

$$\phi_n(t) = \gamma_1 t^n + \gamma_2 t^{n-1} + \dots + \gamma_{n+1} + \frac{\Omega(t)}{t^p}. \quad (4.15)$$

Moreover, since $\Omega(t)$ is B.V., it tends to a limit $\Omega(+\infty)$ as $t \rightarrow +\infty$. Thus $s {}^*\Omega(s) \rightarrow \Omega(+\infty)$ as $s \rightarrow 0+$. But we must have $s {}^*\Omega(s) = o(1)$, for (4.14) to agree with (4.10). Hence $\Omega(+\infty) = 0$; and since the remainder term $t^{-p}\Omega(t)$ must be bounded (by Theorem 1), we can rewrite this remainder term as $\lambda(t)/(1+t)^p$, where $\lambda(t)$ is of B.V. and vanishes at $+\infty$.

We shall next prove: (a) that $\lambda(t)$ satisfies (*) and hence belongs to $\mathcal{B}$; (b) that $\lambda(t)$ belongs to $\mathcal{L}$, and hence to $\mathcal{R}$. It will plainly be enough to prove (a) and (b) for $\Omega(t)$ instead of $\lambda(t)$. Let us, temporarily, assume that we are given only that $\mu_{n+p} < \infty$ and not the full information $\mu_{n+p+1} < \infty$. Then the argument that led to (4.14) is still possible, but we can no longer assume Ω to be B.V. However, reference to the argument following (4.13) shows that ${}^*\Omega$ is the sum of terms of the forms

- (i) $\gamma_{n+p+1} {}^*G_{[\rho]} \quad (\rho \leq n+r+t),$
- (ii) ${}^*G {}^*M_{n+r+t}^{n+r+t} \quad (r+t \leq p).$

All terms of type (i) are L.T.'s of functions of B.V. and must automatically satisfy (**). By Lemma 11 (ii) the terms of type (ii) are also L.T.'s of functions which satisfy (**). Thus Ω satisfies (**). In other words we have shown that if $\mu_{n+p+1} < \infty$ then we may write

$$\phi_n(t) = \gamma_1 t^n + \gamma_2 t^{n-1} + \dots + \gamma_{n+1} + \frac{\Lambda(t)}{t^{1+p}}, \quad (4.16)$$

where $\Lambda(t)$ satisfies (**). If we compare (4.15) with (4.16) we deduce that $\Omega(t) = \Lambda(t)/t$. Hence

$$\Omega(t) - \Omega(t - \alpha) = \frac{\Lambda(t) - \Lambda(t - \alpha)}{t} - \frac{\alpha}{t} \Omega(t - \alpha),$$

and the demonstration that $\Omega(t)$ satisfies (*) follows from the facts that $\Lambda(t)$ satisfies (**) and that $\Omega(t)$ vanishes at $+\infty$. This proves (a).

To prove (b) let us suppose temporarily that we are given that $\mu_{n+p+2} < \infty$. Consider terms of type (i). The distribution function G involved in a typical term was shown above to have its first $n+r+t$ moments finite, when $\mu_{n+p+1} < \infty$. We are now assuming that $\mu_{n+p+2} < \infty$, and so we can take it that G has its first $n+r+t+1$ moments finite. Hence, the first moment of $G_{[\rho]}$ will be finite, and representable as γ_{n+p+2} . By Lemma 4 we can therefore infer that terms of type (i) can always be written as

$$\frac{\gamma_{n+p+1}}{s} + \gamma_{n+p+2} * G_{[\rho+1]}(s). \quad (4.17)$$

In addition, terms of type (ii) can, by Lemma 10 and the finiteness of μ_{n+p+2} , be written as

$$\frac{\gamma_{n+p+2} * G(s)}{s} + *G(s) * M_{n+p+1}^{n+p}(s),$$

$$\text{and thence as} \quad \frac{\gamma_{n+p+2}}{s} + \gamma_{n+p+2} * G_{[1]}(s) + *G(s) * M_{n+p+1}^{n+p}(s). \quad (4.18)$$

If we combine the new expressions (4.17) and (4.18) of all terms of types (i) and (ii) then it is evident that $*\Omega$ can be expressed in the form

$$\frac{\gamma_{n+p+2}}{s} + *\Xi(s),$$

where $\Xi(t)$ is some function of B.V. But a (necessary and) sufficient condition for a function to belong to $L_1(0, \infty)$ is that its L.T. be recognizable as the L.-S.T. of a function of B.V. Hence $\Omega(t) - \gamma_{n+p+2}$ belongs to L_1 . But $\Omega(t)$ vanishes at $+\infty$. Thus $\Omega(t)$ belongs to L_1 . What we have hereby succeeded in proving may now be stated as follows: if $\mu_{n+p+1} < \infty$ then

$$\phi_n(t) = \gamma_1 t^n + \gamma_2 t^{n-1} + \dots + \gamma_{n+1} + \frac{\Pi(t)}{t^{p-1}}, \quad (4.19)$$

where $\Pi(t) \in L_1$. If we compare (4.19) with (4.15) we deduce that $\Pi(t) = \Omega(t)/t$. From this equation the proof of (b) is apparent. This brings to an end the proof of Theorem 2.

We have now proved the required polynomial approximations to the ϕ -moments, and are ready to turn, in the next section, to the study of the ψ -cumulants. Before proceeding, however, let us see what can be said about $\phi_n(t)$ if $\mu_{n+1} = \infty$. In particular, suppose we are only given that $\mu_{q+1} < \infty$, $q < n$. Then, as in the proof of Theorem 1, we can at least get as far as

$$*\phi_n(s) = \frac{\gamma_1}{s^n} + \frac{\gamma_2}{s^{n-1}} + \dots + \frac{\gamma_q}{s^{n-q+1}} + \frac{*M_q^q}{s^{n-q}} + \dots + *M_q^n.$$

The following is then an easy deduction from Lemma 12.

COROLLARY 1. If $\mu_{q+1} < \infty$, $q > 0$, then

$$\phi_n(t) = \gamma_1 t^n + \gamma_2 t^{n-1} + \dots + \gamma_{q+1} t^{n-q} + (1+t)^{n-q} \lambda(t),$$

where $\lambda(t) \in \mathcal{B}$.

Our results would have greater symmetry if we were able to replace $\mathcal{B}$ by $\mathcal{R}$ in Corollary 1, but we have been unable to prove that this replacement is allowable.

5. ASYMPTOTIC BEHAVIOUR OF THE ψ -CUMULANTS

THEOREM 3. If $\mu_{n+p+1} < \infty$, $p > 0$, then

$$\psi_n(t) = \gamma_n t + \gamma_{n+1} + \frac{\lambda(t)}{(1+t)^p},$$

where $\lambda(t) \in \mathcal{R}$.

Proof. $\psi_n(t)$ is a linear combination, with numerical weights, of a finite number of terms like

$$\phi_{p_1}(t) \phi_{p_2}(t) \dots \phi_{p_r}(t),$$

where $p_1 + p_2 + \dots + p_r = n$. By Theorem 2 we have

$$\phi_{p_j}(t) = \gamma_1 t^{p_j} + \gamma_2 t^{p_j-1} + \dots + \gamma_{p_j+1} + \frac{\lambda_{p_j}(t)}{(1+t)^{n+p-p_j}},$$

where $\lambda_{p_j}(t) \in \mathcal{R}$. Hence

$$\phi_{p_1}(t) \phi_{p_2}(t) \dots \phi_{p_r}(t) = \gamma_1 t^n + \gamma_2 t^{n-1} + \dots + \gamma_n + R(t),$$

where $R(t)$ is the sum of a finite number of terms of the form

$$t^m \times \frac{\lambda_{l_1}(t)}{(1+t)^{n+p-l_1}} \times \frac{\lambda_{l_2}(t)}{(1+t)^{n+p-l_2}} \times \dots \times \frac{\lambda_{l_s}(t)}{(1+t)^{n+p-l_s}},$$

with $m + l_1 + l_2 + \dots + l_s \leq n$. By property (C5) of the class $\mathcal{R}$, this form can be rewritten

$$\frac{\lambda(t)}{(1+t)^p} \left\{ \frac{t^m}{(1+t)^{sn+(s-1)p-\sum l_i}} \right\},$$

where $\lambda(t) \in \mathcal{R}$. But we must have $sn + (s-1)p - \sum l_i \geq m$, so the function in the braces is clearly in the class $\mathcal{R}$. A second appeal to (C5) allows us to write

$$R(t) = \frac{\lambda(t)}{(1+t)^p}$$

for some $\lambda(t) \in \mathcal{R}$. Combining all such remainder terms we discover that

$$\psi_n(t) = \gamma_1 t^n + \gamma_2 t^{n-1} + \dots + \gamma_{n+1} + \frac{\lambda(t)}{(1+t)^p} \quad (\lambda \in \mathcal{R}). \quad (5.1)$$

The theorem will be proved if we can show that the coefficients $\gamma_1, \gamma_2, \dots, \gamma_{n-1}$ in the right-hand side of (5.1) all vanish. These coefficients depend on the first $n-1$ moments $\mu_1, \mu_2, \dots, \mu_{n-1}$ of $F(x)$ and on no other moments. It might be possible to follow through the algebra of their derivation and then to show that they automatically vanish, but the algebra involved seems prohibitive (although we return to this topic in the appendix). For the present proof we adopt a different attack.

Let $\mathcal{F}_{n-1}$ be the space of all distribution functions of non-negative random variables whose first $(n-1)$ moments are finite, and let $\mathcal{L}_{n-1}$ be the Euclidean space of $(n-1)$ dimensions. Then we can map $\mathcal{F}_{n-1}$ into $\mathcal{L}_{n-1}$ by letting $F \in \mathcal{F}_{n-1}$ map into the point $\mu = (\mu_1, \mu_2, \dots, \mu_{n-1})$ whose co-ordinates are the first $(n-1)$ moments, in the correct order, of F . Call this mapping T_{n-1} so that $T_{n-1}F = \mu$. Then we wish to show that the rational functions $\gamma_1, \gamma_2, \dots, \gamma_{n-1}$ of $(\mu_1, \mu_2, \dots, \mu_{n-1})$ vanish throughout $T_{n-1}\mathcal{F}_{n-1}$. It will be sufficient if we can find an $(n-1)$ -sphere $S_{(n-1)} \subset T_{n-1}\mathcal{F}_{n-1}$ and show that these rational functions vanish throughout $S_{(n-1)}$. Let $\mathcal{G}_{n-1}$ be the class of $F(x)$ which can be written

$$1 - \frac{1}{(n-1)} \{e^{-x/\nu_1} + e^{-x/\nu_2} + \dots + e^{-x/\nu_{n-1}}\} \quad (x \geq 0),$$

where $\nu_1, \nu_2, \dots, \nu_{n-1}$ are real and positive. We show below that $T_{n-1} \mathcal{G}_{n-1}$ contains an $S_{(n-1)}$. Thus it will, in fact, be sufficient to prove that $\gamma_1, \gamma_2, \dots, \gamma_{n-1}$ vanish whenever $F \in \mathcal{G}_{n-1}$.

To prove that $T_{n-1} \mathcal{G}_{n-1}$ contains an $S_{(n-1)}$ we remark that if $F \in \mathcal{G}_{n-1}$ then the $\{\mu_r\}$ are given as elementary functions of $\nu_1, \nu_2, \dots, \nu_{n-1}$:

$$\mu_r = \frac{r!}{(n-1)} \{\nu_1^r + \nu_2^r + \dots + \nu_{n-1}^r\} \quad (r = 1, 2, \dots).$$

It is simple to show that there is a point $(\nu_1, \nu_2, \dots, \nu_{n-1})$ in a spherical neighbourhood of which the Jacobian

$$\frac{\partial(\mu_1, \mu_2, \dots, \mu_{n-1})}{\partial(\nu_1, \nu_2, \dots, \nu_{n-1})}$$

does not vanish. Thus there is a point $(\mu_1, \mu_2, \dots, \mu_{n-1})$, a spherical neighbourhood of which is in $T_{n-1} \mathcal{G}_{n-1}$.

Suppose $F \in \mathcal{G}_{n-1}$ and generates a renewal process $X_1, X_2, \dots$, in the usual way. Let us suppose $\nu_1 \leq \nu_i$, all i , and write

$$F^\#(x) = 1 - e^{-x/\nu_1}.$$

Then we can uniquely determine a renewal process $X_1^\#, X_2^\#, \dots$ by demanding $F^\#(X_i^\#) = F(X_i)$, all i . By the construction of $F^\#$ we see that $X_i^\# \leq X_i$, all i , and hence, if we employ obvious extensions of our notation, $N_i^\# \geq N_i$. This last inequality implies $\phi_n^\#(t) \geq \phi_n(t)$ for all n , and so the generating function $\Phi_i(\zeta)$ is undoubtedly analytic in any circle $C_\epsilon \equiv \{\zeta: |\zeta| < \epsilon\}$ within which $\Phi_i^\#(\zeta)$ is analytic. But a routine calculation shows that

$$\Phi_i^\#(\zeta) = \frac{\exp\{\zeta/\nu_1(1-\zeta)\}}{(1-\zeta)},$$

which is analytic in C_1 . Thus $\Phi_i(\zeta)$ is analytic in C_1 . We use this fact presently.

If $F \in \mathcal{G}_{n-1}$ then

$$*F(s) = \frac{1}{n-1} \sum_1^{n-1} \frac{1}{1+\nu_i s}. \quad (5.2)$$

Let $s = \sigma + i\tau$ with σ, τ real. Then it is evident from (5.2) that $*F(\sigma)$ is real and strictly decreasing in $I \equiv [0, +\infty)$, and that if $\sigma \in I$ and $\tau \neq 0$ then $|*F(s)| < *F(\sigma)$. Thus the $(n-1)$ zeros of $*F(s) - 1$ must consist of one zero at $s = 0$ and $(n-2)$ zeros in the open half-plane $\sigma < 0$. Furthermore, all these $(n-1)$ zeros must be located at regular points of $*F(s)$, although some may be located at zeros of $*F'(s)$ so that the possibility of multiple zeros cannot be ruled out. Now consider the equation

$$*F(s) = 1 - \zeta, \quad (5.3)$$

where ζ is complex, and confined to some small neighbourhood, C_ϵ , of the origin. Then it follows from the previous discussion of this paragraph, and from the theory of implicit functions (Goursat (1933), vol. II, p. 287) that (5.3) has $(n-1)$ roots $z_1(\zeta), z_2(\zeta), \dots, z_{n-1}(\zeta)$ which are continuous functions of ζ . With no loss of generality we can take it that $z_1(0) = 0$. Then, if ϵ, δ are sufficiently small, we shall have $z_2(\zeta), z_3(\zeta), \dots, z_{n-1}(\zeta)$ confined to the open half-plane $\sigma < -2\delta$, $|z_1(\zeta)| < \delta$, whenever $\zeta \in C_\epsilon$. However, because of the possibility of multiple zeros of $*F(s) - 1$ we cannot be sure that $z_2(\zeta), z_3(\zeta), \dots, z_{n-1}(\zeta)$ always have distinct values unless we restrict ζ to an annulus $A_\epsilon \equiv \{\zeta: \frac{1}{2}\epsilon < |\zeta| < \epsilon\}$.

It will be convenient to write $*F(s) = Q(s)/P(s)$, where $Q(s)$ is a polynomial of degree $(n-2)$, $P(s)$ a polynomial of degree $(n-1)$. This is legitimate, by (5.2). If we write

$$*\Phi_s(\zeta) = \int_0^\infty e^{-st} \Phi_t(\zeta) dt,$$

then Lemma 6 leads to the conclusion

$$\begin{aligned} {}^*\Phi_s(\zeta) &= \frac{1}{s} [1 - \zeta \{1 - {}^*F(s)\}^{-1}]^{-1} \\ &= \frac{1}{s} \frac{P(s) - Q(s)}{(1 - \zeta) P(s) - Q(s)}. \end{aligned} \quad (5.4)$$

Apart from the factor s , the polynomial in the denominator of (5.4) evidently vanishes at $z_1(\zeta), z_2(\zeta), \dots, z_{n-1}(\zeta)$. Bearing in mind that $z_1(0) = 0$, some elementary manipulation shows that (5.4) may be replaced by

$${}^*\Phi_s(\zeta) = - \frac{\prod_{i=2}^{n-1} [s - z_i(0)]}{\prod_{i=1}^{n-1} [s - z_i(\zeta)]}. \quad (5.5)$$

Thus we may adopt a partial fraction expansion

$${}^*\Phi_s(\zeta) = \sum_{i=1}^{n-1} \frac{Z_i(\zeta)}{s - z_i(\zeta)}, \quad (5.6)$$

where

$$Z_j(\zeta) = \frac{\prod_{i=2}^{n-1} [z_j(\zeta) - z_i(0)]}{\prod_{\substack{i=1 \\ i \neq j}}^{n-1} [z_j(\zeta) - z_i(\zeta)]} \quad (j = 1, 2, \dots, n-1). \quad (5.7)$$

Since, for $\zeta \in A_e$, the $z_i(\zeta)$ are distinct-valued continuous functions of ζ it follows from (5.7) that the $Z_j(\zeta)$ are finite-valued, continuous functions of $\zeta \in A_e$. Thus we can find a bound $M < \infty$ such that $|z_j(\zeta)| < M$ for all j , uniformly in $\zeta \in A_e$. Moreover, it is apparent that $Z_1(\zeta)$ cannot vanish in A_e for this would require that one of the factors in the numerator of (5.7) vanishes in A_e . But if $\zeta \in A_e$ then $|z_1(\zeta)| < \delta$, and $|z_j(0)| > 2\delta$ for all $j \neq 1$. Thus $Z_1(\zeta)$ cannot vanish in A_e and so we can find a bound $m > 0$ such that $|Z_1(\zeta)| > m$ for all $\zeta \in A_e$.

We next invert the L.T. (5.6) and obtain

$$\Phi_t(\zeta) = \sum_{i=1}^{n-1} Z_i(\zeta) e^{tz_i(\zeta)}. \quad (5.8)$$

If we suppose $\zeta \in A_e$ and use the results of the preceding paragraph then (5.8) can be rewritten

$$\Phi_t(\zeta) = Z_1(\zeta) e^{tz_1(\zeta)} \{1 + S_1(\zeta, t)\},$$

where $S_1(\zeta, t) = 0(e^{-\delta t})$ uniformly in ζ . If we take logarithms we have

$$\Psi_t(\zeta) = tz_1(\zeta) + \log Z_1(\zeta) + S_2(\zeta, t), \quad (5.9)$$

where $S_2(\zeta, t)$ is also $0(e^{-\delta t})$ uniformly in $\zeta \in A_e$.

We have shown earlier that $\Phi_t(\zeta)$ is analytic in C_1 and it is plain from the definition of $\Phi_t(\zeta)$ that $\Phi_t(0) = 1$. Thus, in some C_e , $\Psi_t(\zeta)$ is an analytic function of ζ . Hence

$$\psi_n(t) = \frac{n!}{2\pi i} \oint_{C'} \frac{\Psi_t(\zeta)}{\zeta^{n+1}} d\zeta,$$

where $C' = C_{\frac{3}{4}e}$. If we now make use of (5.9) we can obtain the result

$$\psi_n(t) = \alpha_n t + \beta_n + 0(e^{-\delta t}), \quad (5.10)$$

by the definition of $z_1(s)$; in other words we have s as a Taylor series in z_1 .† Our problem is therefore merely one of inverting a series. This process has been discussed by many authors, and the fullest computations appear to be those given by Van Orstrand (1910), but one may also refer to Adams & Hippisley (1922), Dwight (1957), and Goursat (1933, vol. I, p. 474). If we use Van Orstrand's results we can obtain the values of α_n listed below. In point of fact our values of α_n were calculated independently, by a method which we describe in § 7 and depends on the correctness of the calculations for β_n , and then compared with Van Orstrand's results. This procedure provided an additional check on our β_n calculations. In quoting our results we take $\mu_1 = 1$. We found:

$$\alpha_2 = \mu_2,$$

$$\alpha_3 = -\mu_3 + 3\mu_2^2,$$

$$\alpha_4 = \mu_4 - 10\mu_2\mu_3 + 15\mu_2^3,$$

$$\alpha_5 = -\mu_5 + 15\mu_2\mu_4 + 10\mu_3^2 - 105\mu_2^2\mu_3 + 105\mu_2^4,$$

$$\alpha_6 = \mu_6 - 21\mu_2\mu_5 - 35\mu_3\mu_4 + 210\mu_2^2\mu_4 + 280\mu_2\mu_3^2 - 1260\mu_2^3\mu_3 + 945\mu_2^5,$$

$$\alpha_7 = -\mu_7 + 28\mu_2\mu_6 + 56\mu_3\mu_5 - 378\mu_2^2\mu_5 + 35\mu_4^2 - 1260\mu_2\mu_3\mu_4 + 3150\mu_2^3\mu_4 - 280\mu_3^3 + 6300\mu_2^2\mu_3^2 - 17,325\mu_2^4\mu_3 + 10,395\mu_2^6,$$

$$\alpha_8 = \mu_8 - 36\mu_2\mu_7 - 84\mu_3\mu_6 + 630\mu_2^2\mu_6 - 126\mu_4\mu_5 + 2520\mu_2\mu_3\mu_5 - 6930\mu_2^3\mu_5 + 1575\mu_2\mu_4^2 + 2100\mu_3^2\mu_4 - 34,650\mu_2^2\mu_3\mu_4 + 51,975\mu_2^4\mu_4 - 15,400\mu_2\mu_3^3 + 138,600\mu_2^3\mu_3^2 - 270,270\mu_2^5\mu_3 + 135,135\mu_2^7.$$

Let us now turn to the more involved problem of calculating β_n . Lemma 4 tells us there is a distribution function $F_{(1)}$ such that

$$*F_{(1)}(s) = \frac{1 - *F(s)}{\mu_1 s},$$

and if, for typographical simplicity, we write $\tau_1, \tau_2, \dots$ for the moments of $F_{(1)}$, then

$$*F_{(1)}(s) = \sum_{r=0}^{\infty} \frac{\tau_r (-s)^r}{r!}. \quad (6.2)$$

By Lemma 6 and (6.2) we have therefore that

$$\frac{* \phi_n(s)}{n!} = \frac{1}{s^n} \left\{ 1 - \tau_1 s + \frac{\tau_2}{2!} s^2 - \dots \right\}^{-n},$$

and the coefficient of s^0 in the Laurent expansion of the right-hand side of this equation is the coefficient of s^n in

$$1 + n \left[\tau_1 s - \frac{\tau_2}{2!} s^2 + \dots \right] + \frac{n(n+1)}{2!} \left[\tau_1 s - \frac{\tau_2}{2!} s^2 + \dots \right]^2 + \frac{n(n+1)(n+2)}{3!} \left[\tau_1 s - \frac{\tau_2}{2!} s^2 + \dots \right]^3 + \text{etc.},$$

which is evidently

$$\sigma_n = (-1)^n \sum_{d=1}^{d=n} \left\{ (-1)^d n(n+1) \dots (n+d-1) \Sigma^* \frac{\tau_{\pi_1}^{\rho_1} \tau_{\pi_2}^{\rho_2} \dots \tau_{\pi_\nu}^{\rho_\nu}}{\rho_1! \rho_2! \dots \rho_\nu! (\pi_1!)^{\rho_1} (\pi_2!)^{\rho_2} \dots (\pi_\nu!)^{\rho_\nu}} \right\}, \quad (6.3)$$

where the inner Σ^* is a summation over all positive integers $\pi_1, \pi_2, \dots, \pi_\nu; \rho_1, \rho_2, \dots, \rho_\nu$ such that the π_i are all different, $\Sigma \rho \pi = n$ and $\Sigma \rho = d$.

† For a heuristic discussion which shows the present significance of the root $z_1(s)$, see Bartlett (1949).

Suppose a function $Q(t)$, say, can be written as $P(t) + r(t)$, where $P(t)$ is a polynomial in t and $r(t)$ tends to zero, as t tends to $+\infty$, faster than any power of t . Let us agree to mean by 'the constant term in $Q(t)$ ' the coefficient of t^0 in $P(t)$ and by the 'linear coefficient in $Q(t)$ ' let us mean the coefficient of t in $P(t)$. When $F \in \mathcal{G}$ all the moments of F are finite and we can therefore speak of the constant term in $\phi_n(t)$, by Theorem 2. But the constant term in $\phi_n(t)$ must be the coefficient of s^0 in the Laurent expansion of ${}^*\phi_n(s)$. Thus the constant term in $\phi_n(t)/n!$ is σ_n , given in (6.3).

It is trivial to verify that the constant term in, say, $\phi_p(t)\phi_q(t)$ is merely the product of the constant term in $\phi_p(t)$ and the constant term in $\phi_q(t)$, and it is equally trivial to extend this result to products of several ϕ_n 's and thence to linear combinations of products of several ϕ_n 's. But, by the well-known relation between moments and cumulants (Kendall & Stuart (1958)),

$$\frac{\psi_n}{n!} = \sum_{d=1}^{d=n} \left\{ (-1)^{d-1} (d-1)! \Sigma^* \frac{\phi_{\pi_1}^{\rho_1} \phi_{\pi_2}^{\rho_2} \dots \phi_{\pi_v}^{\rho_v}}{\rho_1! \rho_2! \dots \rho_v! (\pi_1!)^{\rho_1} (\pi_2!)^{\rho_2} \dots (\pi_v!)^{\rho_v}} \right\}, \quad (6.4)$$

and so we deduce that β_n , the constant term in $\psi_n(t)$, is given by

$$\frac{\beta_n}{n!} = \sum_{d=1}^{d=n} \left\{ (-1)^{d-1} (d-1)! \Sigma^* \frac{\sigma_{\pi_1}^{\rho_1} \sigma_{\pi_2}^{\rho_2} \dots \sigma_{\pi_\nu}^{\rho_\nu}}{\rho_1! \rho_2! \dots \rho_\nu!} \right\}. \quad (6.5)$$

We shall make use of the customary terminology: *weight* and *degree*. For instance, the τ -product $\tau_1^2 \tau_3 \tau_5^3$ is of degree $2 + 1 + 3 = 6$ and weight $2 \times 1 + 1 \times 3 + 3 \times 5 = 20$. It is to be noticed from (6.3) that σ_n is a certain linear combination of all τ -products of weight n . Similarly, from (6.5), we see that β_n is a certain linear combination of all σ -products of weight n . Hence β_n is a certain linear combination of all τ -products of weight n , in turn, and our method of computing β_n will involve taking each τ -product of weight n , in turn, and determining its coefficient in β_n . (It is worth remarking that this is another place where it has been to our advantage to be dealing with ψ -cumulants instead of conventional ones. As can be seen, now, from (2.6) the constant term b_n in the conventional cumulant $k_n(t)$ involves all τ -products of weight $\leq n$.)

To be specific for a moment, consider a term $\tau_1\tau_2^2$ of weight 5 and so occurring in β_5 . How can $\tau_1\tau_2^2$ arise from (6.5)? In various ways. We may obtain $\tau_1\tau_2^2$ in the expansions of σ_5 , $\sigma_1\sigma_4$, $\sigma_2\sigma_3$, or $\sigma_1\sigma_2^2$. Thus, if we represent $\tau_1\tau_2^2$ by the multipartite number $(1, 2^2)$ the various ways in which $\tau_1\tau_2^2$ will arise in (6.5) correspond to the various separations $(1, 2^2)$, $(1)(2^2)$, $(2)(1, 2)$, $(1)(2)^2$. For each separation we must calculate the individual contribution to the coefficient of $\tau_1\tau_2^2$ in β_5 .

Suppose we now consider a general separation

$$\Pi \equiv S_{11}^{r_{11}} S_{12}^{r_{12}} \dots S_{1a}^{r_{1a}} S_{21}^{r_{21}} \dots S_{2b}^{r_{2b}} \dots S_{kf}^{r_{kf}}$$

$\Pi \equiv S_{11}^{r_{11}} S_{12}^{r_{12}} \dots S_{1a}^{r_{1a}} S_{21}^{r_{21}} S_{22}^{r_{22}} \dots S_{2b}^{r_{2b}} \dots$ of some general multipartite number S . In this $S_{11}, S_{12}, \dots, S_{1a}$ are a different multipartite numbers of the same weight w_1 ; $S_{21}, S_{22}, \dots, S_{2b}$ are b different multipartite numbers of weight w_2 ; and so on. Then the weight of S is $w = w_1 r_1 + w_2 r_2 + \dots + w_k r_k$, where $r_1 = r_{11} + r_{12} + \dots + r_{1a}$, etc. This general separation represents one set of ways in which the τ -product T corresponding to S could arise in (6.5). In fact, the term in (6.5) which is involved is

(6.6)

$$\sigma_{w_1}^{r_1} \sigma_{w_2}^{r_2} \dots \sigma_{w_k}^{r_k}.$$

If we write $r = r_1 + r_2 + \dots + r_k$ then it is evident that the coefficient of (6.6) in (6.5) is

$$\frac{(-1)^{r-1} (r-1)!}{r_1! r_2! \dots r_k!}.$$

(6.7)

(6-7)

B. Research in Psychol. Research

3.3.81

5644

Now consider the term $\sigma_{w_1}^{r_1}$ which must give rise to a τ -product corresponding to the separation $S_{11}^{r_1} S_{12}^{r_2} \dots S_{1a}^{r_a}$. By the multinomial theorem this τ -product can arise in

$$\frac{r_1!}{r_{11}! r_{12}! \dots r_{1a}!}$$

ways. Similarly for the τ -products arising from $\sigma_{w_2}^{r_2}, \dots, \sigma_{w_k}^{r_k}$. Thus the general separation Π arises from the term (6.6) in

$$\frac{r_1! r_2! \dots r_k!}{r_{11}! r_{12}! \dots r_{1a}! r_{21}! r_{22}! \dots r_{2b}! \dots r_{kf}!} \quad (6.8)$$

ways. If we combine the coefficient (6.7), which must multiply every τ -product from (6.6), with the number of ways (6.8) in which T may arise from (6.6) then we derive the *form factor*

$$\frac{(-1)^{r-1} (r-1)!}{r_{11}! r_{12}! \dots r_{kf}!}. \quad (6.9)$$

Notice that the form factor (6.9) depends only on the general 'shape' of the separation Π and not on the weights of the individual separates. Thus, for example, $(1, 2^2, 3) (2)^2 (4)$ and $(1) (3)^2 (7)$ have the same 'shape' which we will denote $(\cdot)(\cdot)(\cdot)^2$. We can thus construct a table of form factors which may be used for a variety of different separations.

The form factor (6.9) is merely the number of ways in which T can arise from (6.6), multiplied by the number (6.7) which must be a factor of every contribution from (6.6), to the final coefficient in β_w . Each separate S_{ij} in Π will also contribute a *numerical factor* which we now deduce from (6.3). Suppose $S_{ij} \equiv (\pi_1^{\rho_1} \pi_2^{\rho_2} \dots \pi_l^{\rho_l})$, corresponding to the τ -product

$$T_{ij} \equiv \tau_{\pi_1}^{\rho_1} \tau_{\pi_2}^{\rho_2} \dots \tau_{\pi_l}^{\rho_l}$$

with

$$w_i = \rho_1 \pi_1 + \rho_2 \pi_2 + \dots + \rho_l \pi_l,$$

$$d_{ij} = \rho_1 + \rho_2 + \dots + \rho_l.$$

Then the coefficient of T_{ij} in (6.3) is evidently

$$\frac{(-1)^{w_i+d_{ij}} w_i (w_i + 1) \dots (w_i + d_{ij} - 1)}{\rho_1! \rho_2! \dots \rho_v! (\pi_1!)^{\rho_1} \dots (\pi_v!)^{\rho_v}}. \quad (6.10)$$

Let us write

$$c_{ij} = \frac{w_i (w_i + 1) \dots (w_i + d_{ij} - 1)}{\rho_1! \rho_2! \dots \rho_v!}$$

for the *separate factor* of S_{ij} . Then the numerical factor (6.10) may be rewritten

$$(-1)^{w_i+d_{ij}} \frac{c_{ij}}{(\pi_1!)^{\rho_1} \dots (\pi_v!)^{\rho_v}}. \quad (6.11)$$

Suppose now that

$$T \equiv \tau_{u_1}^{v_1} \tau_{u_2}^{v_2} \dots \tau_{u_l}^{v_l}$$

is of weight w and degree d . Then when we multiply the numerical factors (6.11) together, one for each separate S_{ij} , we plainly must end with

$$\left\{ \frac{(-1)^{w+d}}{(u_1!)^{v_1} (u_2!)^{v_2} \dots (u_l!)^{v_l}} \right\} \prod_{i,j} c_{ij}. \quad (6.12)$$

In the resultant numerical multiplier (6.12) it may be seen that the factor in braces depends only on T and not on the particular separation Π involved. We may call this a τ -factor.

Thus to determine the coefficient of T in $\beta_w/w!$ we first determine its τ -factor. Then represent T by the appropriate multipartite number S and write down every possible separation

of S . For each separation we multiply together the *separate factors* of each separate, and multiply the product by the *form factor* of the separation. The resulting products for each separation of S are then summed, and the total multiplied by the τ -factor of T . The result is the coefficient of T in $\beta_w/w!$. To convert the result to an expression involving μ 's instead of τ 's we use the result

$$\tau_r = \frac{\mu_{r+1}}{(r+1)\mu_1}.$$

Example. Find the coefficient of $\tau_1\tau_2^2$ in $\beta_5/5!$.

(i) The τ -factor is evidently $\frac{(-)^{5+3}}{(2!)^2} = \frac{1}{4}$.

(ii) $S \equiv (1, 2^2)$ and the possible separations of S are

$$\begin{aligned}\Pi_1 &\equiv (1, 2^2); & \Pi_3 &\equiv (1, 2)(2); \\ \Pi_2 &\equiv (1)(2^2); & \Pi_4 &\equiv (1)(2)^2.\end{aligned}$$

We need the following short tables.

Separate	Factor
(1)	1
(2)	2
(1, 2)	12
(2 ²)	10
(1, 2 ²)	105

Form	Factor
(.)	+1
(.)(.)	-1
(.)(.) ²	+1

Using the tabulated separate and form factors we can compute the required coefficient as follows.

Separation	Form factor	Separation factor	Product
Π_1	+1	105	+ 105
Π_2	-1	1×10	- 10
Π_3	-1	12×2	- 24
Π_4	+1	1×2^2	+ 4
Total			+ 75

If we multiply the total by the τ -factor $\frac{1}{4}$ we find that $\beta_5/5!$ involves a term $75\tau_1\tau_2^2/4$. It is an easy deduction from this result that β_5 involves a term

$$\frac{75 \times 5!}{4 \times 2 \times 3^2} \mu_2 \mu_3^2 = 125 \mu_2 \mu_3^2,$$

(we have taken $\mu_1 = 1$).

For this example we have purposely chosen a fairly simple case, but the calculations become increasingly tedious as we deal with β 's of higher weights. The chief source of error to be guarded against is that some separations of a multipartite number may be overlooked.

For instance, the multipartite number $(1^5, 2)$ referring to the term $\mu_2^5 \mu_3$ in β_7 has 19 different separations.

By the method outlined above we have computed formulae for $\beta_1, \beta_2, \dots, \beta_8$. We now list our results, taking $\mu_1 = 1$.

$$\beta_1 = \frac{1}{2}\mu_2,$$

$$\beta_2 = -\frac{2}{3}\mu_3 + \frac{5}{4}\mu_2^2,$$

$$\beta_3 = \frac{3}{4}\mu_4 - 5\mu_2\mu_3 + \frac{11}{2}\mu_2^3,$$

$$\beta_4 = -\frac{4}{5}\mu_5 + \frac{17}{2}\mu_2\mu_4 + \frac{16}{3}\mu_3^2 - 44\mu_2^2\mu_3 + \frac{279}{8}\mu_2^4,$$

$$\beta_5 = \frac{5}{6}\mu_6 - 13\mu_2\mu_5 - 20\mu_3\mu_4 + \frac{395}{4}\mu_2^2\mu_4 + 125\mu_2\mu_3^2 - 465\mu_2^3\mu_3 + \frac{579}{2}\mu_2^5,$$

$$\beta_6 = -\frac{6}{7}\mu_7 + \frac{37}{2}\mu_2\mu_6 + 34\mu_3\mu_5 - 195\mu_2^2\mu_5 + \frac{165}{8}\mu_4^2 - 605\mu_2\mu_3\mu_4 + 1290\mu_2^3\mu_4 - \frac{1160}{9}\mu_3^3 \\ + 2465\mu_2^2\mu_3^2 - 5790\mu_2^4\mu_3 + (11,895/4)\mu_2^6,$$

$$\beta_7 = \frac{7}{8}\mu_8 - 25\mu_2\mu_7 - \frac{161}{3}\mu_3\mu_6 + 350\mu_2^2\mu_6 - 77\mu_4\mu_5 + 1295\mu_2\mu_3\mu_5 - 3108\mu_3^2\mu_5 + \frac{1575}{2}\mu_2\mu_4^2 \\ + 1015\mu_3^2\mu_4 - 14,560\mu_2^2\mu_3\mu_4 + (76,335/4)\mu_2^4\mu_4 - 6230\mu_2\mu_3^3 + 48,860\mu_2^3\mu_3^2 - 83,265\mu_2^5\mu_3 \\ + (72,855/2)\mu_2^7,$$

$$\beta_8 = -\frac{8}{9}\mu_9 + \frac{65}{2}\mu_2\mu_8 + 80\mu_3\mu_7 - 584\mu_2^2\mu_7 + 133\mu_4\mu_6 - 2520\mu_2\mu_3\mu_6 + 6713\mu_2^3\mu_6 + \frac{392}{5}\mu_2^5 \\ - 3626\mu_2\mu_4\mu_5 - 2352\mu_3^2\mu_5 + 37,464\mu_2^2\mu_3\mu_5 - 54,138\mu_2^4\mu_5 - 2870\mu_3\mu_4^2 + (91,315/4)\mu_2^2\mu_4^2 \\ + 59,010\mu_2\mu_3^2\mu_4 - 339,990\mu_2^3\mu_3\mu_4 + (634,935/2)\mu_2^5\mu_4 + (19,040/3)\mu_3^4 - 218,960\mu_2^2\mu_3^3 \\ + 1,019,970\mu_2^4\mu_3^2 - 1,359,960\mu_2^6\mu_3 + (8,294,895/16)\mu_2^8.$$

To conclude this section let us remark, what should be fairly obvious, that one can obtain the $\{a_n\}$ and $\{b_n\}$ of Theorem 0, which refers to conventional cumulants, from the $\{\alpha_n\}$ and $\{\beta_n\}$ by means of the relations (2.6) and (2.7).

7. CHECKS ON THE FORMAL CALCULATIONS

We begin by describing the way in which we obtained the $\{\alpha_n\}$ from the $\{\beta_n\}$ calculations. This method is not the best way of determining the $\{\alpha_n\}$ directly; that developed in § 6, using Van Orstrand's results, is undoubtedly better. But if one has already performed the calculations of $\{\beta_n\}$ then the present method of deriving the $\{\alpha_n\}$ is fairly easy. By comparing these results with those obtainable by the method of § 6 we have a good check that no separations have been overlooked in the $\{\beta_n\}$ computations.

Let ω_n be the linear coefficient in $\phi_n(t)/n!$. Then ω_n will be the coefficient of s^{-1} in the Laurent expansion of $^*\phi_n(s)/n!$. By the same expansion which preceded (6.3) we have therefore that

$$\omega_1 = 1, \\ \omega_{n+1} = (-1)^n \sum_{d=1}^{d=n} \left\{ (-1)^d (n+1)(n+2) \dots (n+d) \Sigma^* \frac{\tau_{\pi_1}^{\rho_1} \tau_{\pi_2}^{\rho_2} \dots \tau_{\pi_\nu}^{\rho_\nu}}{\rho_1! \dots \rho_\nu! (\pi_1!)^{\rho_1} \dots (\pi_\nu!)^{\rho_\nu}} \right\}. \quad (7.1)$$

Thus ω_{n+1} is a linear combination of all τ -products of weight n . Notice, furthermore, that an individual τ -product of degree d in (7.1) has the same coefficient as it would have in σ_n , the constant term of $\phi_n(t)/n!$, except for being multiplied by a 'correction factor' $(n+d)/n$.

It will be easier to understand the argument that follows if we talk about a specific τ -product, so let us consider $\tau_1 \tau_2^2$ (which occurs in $\alpha_6/6!$). Consider, in particular, the separation $(1, 2)(2)$ of the associated multipartite number $(1, 2^2)$. Corresponding to $(1, 2)(2)$ we

have a number of ways in which $\tau_1 \tau_2^2$ can arise in (6.4) (with $n = 6$) in the linear coefficient $\alpha_6/6!$. These ways are:

- (i) $\tau_1 \tau_2$ from ω_4 in $\phi_4(t)/4!$, and τ_2 from σ_2 in $\phi_2(t)/2!$;
- (ii) $\tau_1 \tau_2$ from σ_3 in $\phi_3(t)/3!$, and τ_2 from ω_3 in $\phi_3(t)/3!$;
- (iii) $\tau_1 \tau_2$ from σ_3 in $\phi_3(t)/3!$, τ_2 from σ_2 in $\phi_2(t)/2!$, and the linear coefficient $\omega_1 = 1$ from $\phi_1(t)$.

Let $\mathcal{D}$ be the coefficient of $\tau_1 \tau_2^2$ in $\alpha_6/6!$ and $\mathcal{C}$ the contribution of the separation (1, 2) (2) towards the coefficient of $\tau_1 \tau_2^2$ in $\beta_5/5!$, as described in § 6. Then a careful investigation of (6.4) shows that (i) contributes an amount $(1 + \frac{2}{3})\mathcal{C}$ towards $\mathcal{D}$. The factor $(1 + \frac{2}{3})$ arises as the 'correction factor' of (1, 2). Similarly (ii) contributes an amount $(1 + \frac{1}{2})\mathcal{C}$ towards $\mathcal{D}$.

Suppose Δ is a given 'shape' of p parts (e.g. $(\cdot)(\cdot)^2(\cdot)^3$ and $(\cdot)(\cdot)(\cdot)(\cdot)(\cdot)^2$ are 'shapes' of six parts). Let Δ have a form factor f . Then the augmented 'shape' $(\cdot)\Delta$ has form factor $-pf$. This contention is easily checked by (6.9).

Now consider the contribution from (iii) towards $\mathcal{D}$. Since both $\tau_1 \tau_2$ and τ_2 are here behaving as from constant terms (σ 's) no correction factor is required for them. But $\mathcal{C}$ is based on a form factor of $(\cdot)(\cdot)$, whereas what is required for (iii) is a form factor of $(\cdot)(\cdot)(\cdot)$. By the preceding paragraph we see that (iii) contributes an amount $-2\mathcal{C}$ to $\mathcal{D}$. Hence the total contribution to $\mathcal{D}$ by the separation (1, 2) (2) is $(\frac{2}{3} + \frac{1}{2})\mathcal{C}$.

The preceding discussion gives an indication of the general result which we now describe, and which can be proved by a careful extension of the argument. Let T be a τ -product of weight w , and associated multipartite number S . Let $S_1^{r_1} S_2^{r_2} \dots S_k^{r_k}$ be a separation of S , where the separate S_j has weight w_j and degree d_j . Let $\mathcal{C}$ be the contribution from the separation $S_1^{r_1} S_2^{r_2} \dots S_k^{r_k}$ to the coefficient of T in $\beta_w/w!$. Then the contribution of this separation to the coefficient of T in $\alpha_{w+1}/(w+1)!$ is

$$\left\{ \frac{r_1 d_1}{w_1} + \frac{r_2 d_2}{w_2} + \dots + \frac{r_k d_k}{w_k} \right\} \mathcal{C}. \quad (7.2)$$

It is this result that enables us to obtain the $\{\alpha_n\}$ rapidly from the $\{\beta_n\}$ calculations.

Example. Find the coefficient of $\tau_1 \tau_2^2$ in $\alpha_6/6!$. We borrow the results of the example worked in § 6. In the computation scheme given below, column A is the extreme right-hand column of the computation scheme displayed in the example in § 6; in other words it consists of the various contributions to the coefficient of $\tau_1 \tau_2^2$ in $\beta_5/5!$, but with the τ -factor left out. In the column B below, we give the appropriate value of the expression in braces in (7.2).

Separation	A	B	Product AB
(1, 2 ²)	+ 105	$\frac{2}{3}$	+ 63
(1) (2 ²)	- 10	$1 + \frac{2}{4}$	- 15
(1, 2) (2)	- 24	$\frac{2}{3} + \frac{1}{2}$	- 28
(1) (2) ²	+ 4	$1 + \frac{1}{2}$	+ 8
Total			+ 28

The τ -factor, as before, is $\frac{1}{4}$.
Thus the coefficient of $\tau_1 \tau_2^2$ in $\alpha_6/6!$ is $\frac{1}{4} \times 28 = 7$.

Converting to μ 's instead of τ 's, and taking $\mu_1 = 1$, we see that α_6 contains a term

$$\frac{7 \times 6!}{2 \times 3^2} \mu_2 \mu_3^2 = 280 \mu_2 \mu_3^2.$$

This result will be found to agree with the value of α_6 given in § 6.

The most obvious check to employ is obtained by putting $F(x) = 1 - e^{-x}$, so that $\mu_n = n!$ for all n . Routine computation shows that in this case $\psi_n(t) = (n!)t + (n-1)!$; thus the substitution $\mu_n = n!$ in our formulae for α_n and β_n should give $\alpha_n = n!$ and $\beta_n = (n-1)!$. We call this the Poisson check and all our formulae pass it.

We now derive a useful and wider set of checks. By (5.4) it is evident that

$$* \Phi_s(\zeta) = \frac{1}{s} + \frac{\zeta}{s\{1 - *F(s) - \zeta\}},$$

so that

$$\sum_{n=0}^{\infty} \frac{\sigma_n \zeta^n}{n!} = 1 + \frac{\zeta}{2\pi i} \oint_C \frac{ds}{s\{1 - *F(s) - \zeta\}}, \quad (7.3)$$

where C is some circle $|s| = \delta$ in the s -plane.

Let us suppose, that for some positive integer p ,

$$*F(s) = 1 - s + \frac{(p-1)}{2!} s^2 - \frac{(p-1)(p-2)}{3!} s^3 + \dots + \frac{(-s)^p}{p}. \quad (7.4)$$

It is true that there is no distribution function $F(x)$ with the above expression for its L.-S.T., but we are interested only in formal relations between coefficients so this fact need not trouble us. From (7.4) we see that

$$1 - *F(s) = \frac{(1-s)^p - 1}{-p}$$

and substitution in (7.3) yields

$$\sum_0^{\infty} \frac{\sigma_n \zeta^n}{n!} = 1 - \frac{\zeta p}{2\pi i} \oint_C \frac{ds}{s\{(1-s)^p - 1 + p\zeta\}}. \quad (7.5)$$

For ζ arbitrarily small the contour integral (7.5) has two simple poles within C , at $s = 0$ and at $s = 1 - (1 - p\zeta)^{1/p}$, where the p th root is meant to be the branch taking the value 1 at $\zeta = 0$. A straightforward evaluation of (7.5) by the usual methods of contour integration then gives

$$\sum_0^{\infty} \frac{\sigma_n \zeta^n}{n!} = \frac{\zeta}{\{1 - (1 - p)^{1/p}\} (1 - p\zeta)^{1-(1/p)}}.$$

Thus $\sum_1^{\infty} \frac{\beta_n \zeta^n}{n!} = \log \left(\sum_0^{\infty} \frac{\sigma_n \zeta^n}{n!} \right) = \log \zeta - \log \{1 - (1 - p\zeta)^{1/p}\} - \frac{p-1}{p} \log (1 - p\zeta).$

If we differentiate the last equation with respect to ζ we obtain the more manageable result

$$\begin{aligned} \sum_1^{\infty} \frac{\beta_n \zeta^{n-1}}{(n-1)!} &= \frac{1}{\zeta} - \frac{\{1 - p\zeta\}^{-(p-1)/p}}{1 - (1 - p\zeta)^{1/p}} + \frac{(p-1)}{1 - p\zeta} \\ &= \frac{1}{\zeta} - \frac{1}{p\zeta} [(1 - p\zeta)^{-(p-1)/p} + (1 - p\zeta)^{-(p-2)/p} + \dots + (1 - p\zeta)^{-1/p}] + \frac{(p-1)}{1 - p\zeta}. \end{aligned} \quad (7.6)$$

The right-hand side of (7.6) is in a form suitable for expansion as a power series in ζ , and on comparing coefficients of ζ^{n-1} one deduces that

$$\frac{\beta_n}{(n-1)!} = (p-1)p^{n-1} - \frac{1}{n!p} \sum_{r=1}^{p-1} r(r+p)(r+2p)\dots(r+\overline{n-1}p). \quad (7.7)$$

Referring back to (7.4) it appears therefore that our formulae for β_n should give the values indicated by (7.7) when we substitute

$$\left. \begin{aligned} \mu_1 &= 1, \quad \mu_2 = (p-1), \quad \mu_3 = (p-1)(p-2), \\ \dots \mu_p &= (p-1)!, \quad \mu_n = 0 \quad (n > p). \end{aligned} \right\} \quad (7.8)$$

Evidently, by (7.4) we have

$$z_1(\zeta) = 1 - (1-p\zeta)^{1/p} = \zeta + \frac{(p-1)}{2!} \zeta^2 + \frac{(p-1)(2p-1)}{3!} \zeta^3 + \dots$$

If we pick out the coefficient of ζ^n in this expansion we can deduce that

$$\alpha_n = (p-1)(2p-1)\dots(\overline{n-1}p-1). \quad (7.9)$$

Thus the substitution (7.8) should also give (7.9).

We call the checks just described the p -check and all our quoted formulae have passed p -checks for $2 \leq p \leq 8$. We presume therefore that they are perfectly correct.†

An advantage of the p -check in the computation of the $\{\beta_n\}$ formulae is that the vanishing of the μ_n for $n > p$ enables one to isolate inevitable errors to a few terms. This makes the search for errors easier.

APPENDIX

A combinatorial identity

From Theorem 3 it is possible to deduce a curious combinatorial identity. Let us say that a multipartite number $S \equiv (p_1^{r_1}, p_2^{r_2}, \dots, p_k^{r_k})$ is allowable if its degree d is not less than its weight w (one of the $\{p_j\}$ may be zero). If S is allowable then associate with it a coefficient

$$A(S) = \frac{d!}{(d-w)!(p_1!)^{r_1}(p_2!)^{r_2}\dots(p_k!)^{r_k}}.$$

A separation $S_1^{q_1} S_2^{q_2} \dots S_m^{q_m}$ of S is allowable if every separate S_i is allowable. To an allowable separation associate a coefficient

$$B(S_1^{q_1} S_2^{q_2} \dots S_m^{q_m}) = \frac{(-1)^{q-1}(q-1)! [A(S_1)]^{q_1} \dots [A(S_m)]^{q_m}}{q_1! q_2! \dots q_m!},$$

where $q = q_1 + q_2 + \dots + q_m$.

Then the identity in question is that

$$\sum^{\#} B(S_1^{q_1} S_2^{q_2} \dots S_m^{q_m}) = 0, \quad (A.1)$$

whenever the summation $\sum^{\#}$ is over all allowable separations of an allowable multipartite number S , of degree d , weight w , with $d > w + 1$.

By choosing S in special ways a large number of interesting special cases may be deduced from (A.1) but we shall not pursue this topic. If we were able to prove (A.1) by elementary argument we would be in

† I am indebted to Mr D. A. East of the Statistical Laboratory, University of Cambridge, for the calculation of α_8 , β_8 and the performance of all these checks. Incidentally, I am also indebted to him for noticing an additional check that no separations of a given multipartite number have been omitted in the working. For the form factors summed over all separations of a given multipartite number total to zero, except when the number is of the form $(\pi_1^{\rho_1})$. This can be checked algebraically, for the total involved with a number $(\pi_1^{\rho_1}, \pi_2^{\rho_2}, \dots, \pi_\nu^{\rho_\nu})$ is the coefficient of $x_1^{\rho_1} x_2^{\rho_2} \dots x_\nu^{\rho_\nu}$ in the logarithmic expansion of $\{(1-x_1)(1-x_2)\dots(1-x_\nu)\}^{-1}$.

a position to remove the complex-variable argument needed in the proof of Theorem 3. As it is, we deduce (A. 1) from Theorem 3.

To prove (A.1) suppose that $s^* \phi_1(s) = e_0 + e_1 s + e_2 s^2 + \dots$

Then

$$s^d \frac{d^* \phi_d(s)}{d!} = \sum_{w=0}^{\infty} A_w^d s^w, \quad (\text{A. 2})$$

where

$$A_w^d = d! \Sigma^* \frac{e_{\pi_1}^{\rho_1} e_{\pi_2}^{\rho_2} \dots e_{\pi_v}^{\rho_v}}{(\pi_1!)^{\rho_1} \dots (\pi_v!)^{\rho_v}},$$

and Σ^* has the usual meaning.

If we invert equation (A. 2) we have

$$\frac{\phi_d(t)}{d!} = \sum_{w=0}^{w=d} A_w^d \frac{t^{d-w}}{(d-w)!} + o_d(t), \quad (\text{A. 3})$$

where we may assume (by Theorem 2) that the remainder term $o_d(t)$ is negligible. We deduce from (A. 3) that the coefficient of t^r in $\phi_d(t)$ is a linear combination of e -products of degree d and weight $d-r$. It then follows easily that the same remark must also apply to $\psi_d(t)$.

Consider a particular e -product $E \equiv e_{\sigma_0}^{\sigma_0} e_{\sigma_1}^{\sigma_1} \dots e_{\sigma_m}^{\sigma_m}$, of weight w and degree d , with $\sigma_i \geq 0$ (this is to allow some e_i to be absent). Suppose, moreover, that $d > w + 1$, so that the associated multipartite number $S \equiv (0^{\sigma_0} 1^{\sigma_1} \dots m^{\sigma_m})$ is of the type required for (A. 1) to be true. Then the coefficient of t^{d-w} in $\psi_d(t)$ must contain some multiple of E . But since $d-w > 1$, Theorem 3 insists that the coefficient of t^{d-w} must vanish identically. Hence the multiplier of E must be identically zero. This is, essentially, the proof of (A. 1). We must now effect an alternative evaluation of the multiplier, however, to show why this is so.

Let

$$\Pi \equiv S_{11}^{q_{11}} S_{12}^{q_{12}} \dots S_{1a}^{q_{1a}} S_{21}^{q_{21}} \dots S_{2b}^{q_{2b}} \dots S_{k1}^{q_{k1}} \dots S_{kl}^{q_{kl}}$$

be a separation of S , with S_{ij} of weight w_{ij} and degree d_i for all j . This separation corresponds to a factorization of E into $E_{11}^{q_{11}} E_{12}^{q_{12}} \dots E_{kl}^{q_{kl}}$, where the e -product E_{ij} is that associated with S_{ij} . We are interested in the way E can appear in $\psi_d(t)$. By (A. 3), assuming $o_d(t)$ negligible, it is clear that E_{ij} cannot appear in $\phi_d(t)$ unless S_{ij} is allowable. Thus we must suppose Π to be an allowable separation of S , and it is then apparent that the factorization $E_{11}^{q_{11}} \dots E_{kl}^{q_{kl}}$ must emerge, in $\psi_d(t)$, from the term

$$\frac{(-1)^{q-1} (q-1)! \phi_{d_1}^{q_{11}} \phi_{d_2}^{q_{12}} \dots \phi_{d_k}^{q_{kl}}}{q_1! q_2! \dots q_k! (d_1!)^{q_{11}} \dots (d_k!)^{q_{kl}}}, \quad (\text{A. 4})$$

where

$$q_i = \sum_j q_{ij} \quad \text{and} \quad q = \sum_i q_i.$$

We can obtain $E_{11}^{q_{11}} E_{12}^{q_{12}} \dots E_{1a}^{q_{1a}}$ from $\phi_{d_1}^{q_{11}}$ in

$$\frac{q_1!}{q_{11}! q_{12}! \dots q_{1a}!}$$

ways, each time with a coefficient which is, by (A. 3), simply

$$[A(S_{11})]^{q_{11}} [A(S_{12})]^{q_{12}} \dots [A(S_{1a})]^{q_{1a}}.$$

Similar remarks apply to the other batches of factors of E of equal degree. We can then deduce from (A. 4) that the contribution to the coefficient of E , in $\psi_d(t)/d!$ which corresponds to the separation Π is

$$\frac{(-1)^q (q-1)! [A(S_{11})]^{q_{11}} \dots [A(S_{kl})]^{q_{kl}}}{q_{11}! q_{12}! \dots q_{kl}!} = B(\Pi).$$

Thus our alternative evaluation of the multiplier of E in $\psi_d(t)/d!$ is simply $\Sigma B(\Pi)$, and the identity (A. 1) is proved.

REFERENCES

- ADAMS, E. P. & HIPPISEY, R. L. (1922). *Smithsonian Mathematical Formulae and Tables of Elliptic Functions*, Smithsonian Misc. Coll. **74** (1). Washington D.C., Smithsonian Institute.
- BARTLETT, M. S. (1949). Some evolutionary stochastic processes. *J.R. Statist. Soc. B*, **11**, 211-29.
- BENEŠ, V. E. (1957). On queues with Poisson arrivals. *Ann. Math. Statist.* **28**, 670-7.
- BLACKWELL, D. (1948). A renewal theorem. *Duke Math. J.* **15**, 145-50.
- COX, D. R. (1955). Some statistical methods connected with series of events. *J.R. Statist. Soc. B*, **17**, 129-64.
- COX, D. R. & SMITH, W. L. (1954). On the superposition of renewal processes. *Biometrika*, **41**, 91-9.
- DWIGHT, H. B. (1957). *Tables of Integrals and other Mathematical Data*, 3rd edition. New York: Macmillan.

- FELLER, W. (1941). On the integral equation of renewal theory. *Ann. Math. Statist.* **12**, 243-67.
- FELLER, W. (1949). Fluctuation theory of recurrent events. *Trans. Amer. Math. Soc.* **67**, 98-119.
- GOURSAT, E. (1933). *Cours d'Analyse Mathématique*, 5th edition. Paris: Gauthier-Villars.
- JORDAN, C. (1947). *Calculus of Finite Differences*, 2nd edition. New York: Chelsea.
- KENDALL, M. G. & STUART, A. (1958). *The Advanced Theory of Statistics*, **1**. London: Charles Griffin.
- PALEY, R. E. A. C. & WIENER, N. (1934). *Fourier Transforms in the Complex Domain*, *Amer. Math. Soc. Colloquium Publications*, XIX. New York.
- SKELLAM, J. G. & SHENTON, L. R. (1957). Distributions associated with random walk and recurrent events. *J.R. Statist. Soc. B*, **19**, 64-118.
- SMITH, W. L. (1953). *Transient Queue and Storage Phenomena*. Chapter 7 of Ph.D. Thesis, University of Cambridge.
- SMITH, W. L. (1954). Asymptotic renewal theorems. *Proc. Roy. Soc. Edinb. A*, **64**, 9-48.
- SMITH, W. L. (1955). Regenerative stochastic processes. *Proc. Roy. Soc. A*, **232**, 6-31.
- SMITH, W. L. (1957). On renewal theory, counter problems, and quasi-Poisson processes. *Proc. Camb. Phil. Soc.* **53**, 175-93.
- TAKÁCS, L. (1956). On a probability problem arising in the theory of counters. *Proc. Camb. Phil. Soc.* **52**, 488-98.
- TÄCKLIND, S. (1945). Fourieranalytische Behandlung vom Erneuerungsproblem. *Skand. AktuarTidskr.* 68-105.
- VAN ORSTRAND, C. E. (1910). Reversion of power series. *Phil. Mag.* **19**, 366-76.

A RANDOM WALK IN WHICH THE STEPS OCCUR RANDOMLY IN TIME

BY A. MERCER*

Atomic Weapons Research Establishment, Aldermaston

AND C. S. SMITH

National Coal Board, London

SUMMARY

The steps of a one-dimensional random walk are positive and occur randomly in time at a fixed mean rate. The sizes of the steps are independent and the size of each step has the same given probability distribution. The distribution of the time to reach a fixed barrier is obtained and asymptotic approximations are derived. The results are extended to the case in which the barrier and the random walk process converge at a constant rate.

1. INTRODUCTION

The problem considered in this paper arose from the study of the wear of conveyor belting. Incidents which damage belting are of various degrees of severity and different types of incident occur with different frequencies. From a given distribution of the life of belting, information is required about the relative importance of different types of incident. In particular, it is desirable to know whether the wear caused by nearly continuous abrasion is more or less important than the damage due to the more destructive but less frequent blows.

The model, which has been used, assumes that damage accumulates until the total reaches a certain level, when the belt is considered to be worn out completely. The damage done by each incident may be considered as a step of a random walk process in which the steps occur at a constant mean rate, and the problem is to relate the distribution of the size of steps to the distribution of the time taken to reach a barrier.

2. TIME TO REACH A FIXED BARRIER

Consider the one-dimensional random walk process in which the steps are positive and occur randomly at mean rate m , but the size of each step, x , is not specified. Suppose that the sizes of the steps are independent and the probability density of the step size is $f(x)$. Assume that the process starts from the point $u = 0$ at time $t = 0$ and let $q(u, t)$ denote the probability that the distance travelled at time t is less than or equal to u . Then

$$q(u, t) = e^{-mt} \sum_{r=0}^{\infty} \frac{(mt)^r}{r!} p_r(u), \quad (1)$$

where $p_r(u)$ is the probability that the distance covered by r steps is less than or equal to u . Therefore, if $p(t)$ is the probability density of the time to reach a barrier at a distance h from the initial position,

$$p(t) = \frac{d}{dt} \{1 - q(h, t)\} = me^{-mt} \sum_{r=0}^{\infty} \frac{(mt)^r}{r!} \{p_r(h) - p_{r+1}(h)\} \quad (2)$$

* Formerly at the National Coal Board.

and the k th moment of the time about time zero is

$$\frac{k}{m^k} \sum_{r=0}^{\infty} \frac{(r+k-1)!}{r!} p_r(h). \quad (3)$$

But $p_r(u) = \int_0^u p_{r-1}(x)f(u-x)dx$ and $p_0(u) = 1$ for $u > 0$, so that if the Laplace transform with respect to x of $f(x)$ is denoted by $F(s)$ and the Laplace transform with respect to u of $p_r(u)$ is denoted by $P_r(s)$, then

$$P_r(s) = P_{r-1}(s)F(s) = \dots = P_0(s)\{F(s)\}^r = \frac{1}{s}\{F(s)\}^r. \quad (4)$$

Therefore the Laplace transform with respect to h of the k th moment of the time about time zero is

$$\frac{k}{m^k} \sum_{r=0}^{\infty} \frac{(r+k-1)!}{r!} \frac{1}{s} \{F(s)\}^r = \frac{k!}{m^k s} \{1 - F(s)\}^{-k}. \quad (5)$$

If $g = \int_0^\infty x f(x) dx$ is assumed to exist for all x , then

$$\frac{G(s)}{F(s)} = \int_0^\infty e^{-sx} f(x) dx = \sum_{r=0}^{\infty} \frac{a_r}{r!} s^r, \quad (6)$$

so that from (5) and (6) as $s \rightarrow 0$, the Laplace transforms with respect to h of the first three moments of the time about time zero are

$$\left. \begin{aligned} \frac{1}{m} \left\{ \frac{1}{a_1 s^2} + \frac{a_2}{2a_1^2 s} + o(1) \right\}, \\ \frac{2}{m^2} \left\{ \frac{1}{a_1^2 s^3} + \frac{a_2}{a_1^3 s^2} + \left(\frac{3a_2^2}{4a_1^4} - \frac{a_3}{2a_1^3} \right) \frac{1}{s} + o(1) \right\}, \\ \frac{6}{m^3} \left\{ \frac{1}{a_1^3 s^4} + \frac{3a_2}{2a_1^4 s^3} + \left(\frac{3a_2^2}{2a_1^5} - \frac{a_3}{2a_1^4} \right) \frac{1}{s^2} + \left(\frac{a_4}{8a_1^4} - \frac{a_2 a_3}{a_1^5} + \frac{5a_3^2}{4a_1^6} \right) \frac{1}{s} + o(1) \right\}. \end{aligned} \right\} \quad (7)$$

But it is well known (for example, Widder, 1946) that if $G(s)$ is the Laplace transform of $g(y)$ with respect to y and as $s \rightarrow +0$, $G(s) \sim As^{-k}$ where k is non-negative, then under certain conditions, $g(y) \sim A\{\Gamma(k)\}^{-1}y^{k-1}$ as $y \rightarrow \infty$. Hence equations (7) suggest that as $h \rightarrow \infty$ the mean, variance and third moment about the mean are

$$\left. \begin{aligned} \mu &= \frac{1}{m} \left\{ \frac{h}{a_1} + \frac{a_2}{2a_1^2} + o(1) \right\}, \\ \sigma^2 &= \frac{1}{m^2} \left\{ \frac{a_2 h}{a_1^3} + \frac{5a_2^2}{4a_1^4} - \frac{2a_3}{3a_1^3} + o(1) \right\}, \\ \mu_3 &= \frac{1}{m^3} \left\{ \left(\frac{3a_2^2}{a_1^5} - \frac{a_3}{a_1^4} \right) h + \frac{11a_2^3}{2a_1^6} - \frac{5a_2 a_3}{a_1^5} + \frac{3a_4}{4a_1^4} + o(1) \right\}. \end{aligned} \right\} \quad (8)$$

Alternatively, the $(i+1)$ th step may be considered as the i th renewal of a process and then a rigorous proof of the asymptotic expressions for the moments can be obtained using results in renewal theory (Smith, 1959).

3. EXAMPLES

The following three examples illustrate the preceding results.

(i) Suppose that the size of step has a negative exponential distribution, $f(x) = \beta e^{-\beta x}$. Then

$$F(s) = \int_0^\infty \beta e^{-(s+\beta)x} dx = \frac{\beta}{\beta+s}.$$

Also $p_r(u)$ can be obtained directly as $e^{-\beta u} \sum_{q=r}^{\infty} \frac{(\beta u)^q}{q!}$ and therefore from (2),

$$p(t) = m e^{-(mt+\beta h)} \sum_{r=0}^{\infty} \frac{(mt)^r}{r!} \frac{(\beta h)^r}{r!}. \quad (9)$$

Hence the mean, variance and third moment about the mean of the times taken to reach the barrier are

$$\left. \begin{aligned} \mu &= \frac{1}{m} (1 + \beta h), \\ \sigma^2 &= \frac{1}{m^2} (1 + 2\beta h), \\ \mu_3 &= \frac{2}{m^3} (1 + 3\beta h). \end{aligned} \right\} \quad (10)$$

But $a_r = \int_0^{\infty} \beta x^r e^{-\beta x} dx = \frac{r!}{\beta^r}$ and substituting in (8) we see that the asymptotic results are exact.

(ii) Suppose that the size of step is the sum of two independent variables which are identically distributed according to the negative exponential distribution.

Then

$$f(x) = \beta^2 x e^{-\beta x} \quad \text{and} \quad F(s) = \left(\frac{\beta}{\beta + s} \right)^2.$$

Also $p_r(u)$ can be obtained directly as

$$e^{-\beta u} \sum_{q=2r}^{\infty} \frac{(\beta u)^q}{q!}$$

and therefore from (2)

$$p(t) = m e^{-(mt+\beta h)} \sum_{r=0}^{\infty} \frac{(mt)^r}{r!} \left\{ \frac{(\beta h)^{2r}}{(2r)!} + \frac{(\beta h)^{2r+1}}{(2r+1)!} \right\} \quad (11)$$

so that as $h \rightarrow \infty$,

$$\left. \begin{aligned} \mu &= \frac{1}{m} \left\{ \frac{3}{4} + \frac{\beta h}{2} + 0(e^{-2\beta h}) \right\}, \\ \sigma^2 &= \frac{1}{m^2} \left\{ \frac{13}{16} + \frac{3\beta h}{4} + 0(\beta h e^{-2\beta h}) \right\}, \\ \mu_3 &= \frac{3}{m^3} \left\{ \frac{9}{16} + \frac{5\beta h}{8} + 0(\beta^2 h^2 e^{-2\beta h}) \right\}. \end{aligned} \right\} \quad (12)$$

But $a_r = (r+1)!/\beta^r$ and on substituting in (8) we see that the asymptotic results only differ from the exact results by terms containing negative exponentials.

(iii) Suppose all the steps are of equal size ξ . Then $2mt$ has the χ^2 -distribution with $2n$ degrees of freedom where n is an integer satisfying $n \geq h/\xi > n-1$ and $\mu = n/m$, $\sigma^2 = n/m^2$ and $\mu_3 = 2n/m^3$. On substituting $a_r = \xi^r$ in (8), we see that the asymptotic values of these three moments are

$$\left. \begin{aligned} \mu &= \frac{1}{m} \left(\frac{h}{\xi} + \frac{1}{2} \right), \\ \sigma^2 &= \frac{1}{m^2} \left(\frac{h}{\xi} + \frac{7}{12} \right), \\ \mu_3 &= \frac{2}{m^3} \left(\frac{h}{\xi} + \frac{5}{8} \right). \end{aligned} \right\} \quad (13)$$

4. TIME TO REACH A MOVING BARRIER

Now suppose that the barrier is moving, relative to the random walk process, at a fixed rate λ per unit time towards the origin. Then at time t , the barrier is at a distance $(h - \lambda t)$ from the origin and the process is certain to have terminated by time h/λ . In the application to the life of conveyor belting, the relative movement is due to continuous wear of the belting, which occurs in addition to the random blows. This wear is the limiting case of that in which very small blows occur frequently. Hence, suppose that the steps consist of those whose size has probability density $f(x)$ and occur at mean rate m and steps of size x_1 occurring at mean rate m_1 . Then if $f_1(x)$ is the density function of the step size distribution

$$(m + m_1)f_1(x) = mf(x) + m_1\delta(x - x_1).^*$$

Therefore

$$(m + m_1)\gamma_r = ma_r + m_1x_1^r, \quad (15)$$

where

$$\gamma_r = \int_0^\infty x^r f_1(x) dx$$

and

$$a_r = \int_0^\infty x^r f(x) dx$$

as previously. Now let $m_1 \rightarrow \infty$ and $x_1 \rightarrow 0$ so that $m_1 x_1 \rightarrow \lambda$. Then from (15),

$$(m + m_1)\gamma_1 \rightarrow ma_1 \left(1 + \frac{1}{\rho}\right) \quad (16)$$

and

$$(m + m_1)\gamma_r \rightarrow ma_r \quad \text{for } r \geq 2, \quad (17)$$

where

$$\rho = \frac{ma_1}{\lambda}. \quad (18)$$

Now the asymptotic values of the mean, variance and third moment about the mean of the distribution of the time to reach the barrier when the step-size distribution has density function $f_1(x)$ are given by replacing a_r by γ_r in (8). Although it has not been proved rigorously, these results suggest that if (16), (17) and (18) are then used, the asymptotic values of the mean, variance and third moment about the mean when the barrier and random walk process converge at a constant rate are,

$$\left. \begin{aligned} \mu &= \frac{1}{m} \left(\frac{\rho}{1+\rho} \right)^2 \left\{ \frac{h}{a_1} \left(\frac{1+\rho}{\rho} \right) + \frac{a_2}{2a_1^2} \right\}, \\ \sigma^2 &= \frac{1}{m^2} \left(\frac{\rho}{1+\rho} \right)^4 \left\{ \frac{a_2 h}{a_1^3} \left(\frac{1+\rho}{\rho} \right) + \frac{5a_2^2}{4a_1^4} - \frac{2a_3}{3a_1^3} \left(\frac{1+\rho}{\rho} \right) \right\}, \\ \mu_3 &= \frac{1}{m^3} \left(\frac{\rho}{1+\rho} \right)^6 \left\{ \left[\frac{3a_2^2}{a_1^5} - \frac{a_3}{a_1^4} \left(\frac{1+\rho}{\rho} \right) \right] \left(\frac{1+\rho}{\rho} \right) h + \frac{11a_2^3}{2a_1^6} - \frac{5a_2 a_3}{a_1^5} \left(\frac{1+\rho}{\rho} \right) + \frac{3a_4}{4a_1^4} \left(\frac{1+\rho}{\rho} \right)^2 \right\}. \end{aligned} \right\} \quad (19)$$

5. PRACTICAL APPLICATION

In the previous sections the distribution of the time to reach a barrier has been found. Expressions for the moments of that time have been obtained as functions of the mean frequency of the steps, the moments of the step-size distribution, the distance of the barrier from the origin and the rate at which the barrier and random walk process converge. In practice, the moments of the life distribution are known and the problem is to make some inference about the discrete blows and continuous wear.

[* For definition of $\delta(h)$ see footnote to p. 68. ED.].

If the variation in the size of the blows is negligible, then from (19) the asymptotic expressions for the first three moments are

$$\left. \begin{aligned} \mu &= \frac{1}{m} \left(\frac{\rho}{1+\rho} \right)^2 \left\{ \left(\frac{1+\rho}{\rho} \right) \frac{h}{\xi} + \frac{1}{2} \right\}, \\ \sigma^2 &= \frac{1}{m^2} \left(\frac{\rho}{1+\rho} \right)^4 \left\{ \left(\frac{1+\rho}{\rho} \right) \frac{h}{\xi} + \frac{5}{4} - \frac{2}{3} \left(\frac{1+\rho}{\rho} \right) \right\}, \\ \mu_3 &= \frac{1}{m^3} \left(\frac{\rho}{1+\rho} \right)^6 \left\{ \left(2 - \frac{1}{\rho} \right) \left(\frac{1+\rho}{\rho} \right) \frac{h}{\xi} + \frac{11}{2} - 5 \left(\frac{1+\rho}{\rho} \right) + \frac{3}{4} \left(\frac{1+\rho}{\rho} \right)^2 \right\}, \end{aligned} \right\} \quad (20)$$

where ξ is the size of the discrete blows. Hence $\rho = m\xi/\lambda$ is the relative importance of the discrete blows to that of the continuous wear. This is the parameter in which one is most interested in order to decide whether the surface of the conveyor belting, which is largely affected by the continuous abrasion, or the carcass, which has to stand heavy infrequent blows, is in greater need of improvement. The exact expressions for the moments are too

Table 1. *Exact and asymptotic values of moments*

ρ	$m\mu$		$m^2\sigma^2$		$m^3\mu_3$	
	Exact	Asymptotic	Exact	Asymptotic	Exact	Asymptotic
0.5	1.056	1.056	0.102	0.102	-0.004	-0.004
1.0	1.625	1.625	0.370	0.370	0.075	0.070
2.0	2.229	2.222	0.901	0.938	0.620	0.565
10.0	2.997	3.140	2.956	2.607	5.455	4.052
∞	3.000	3.500	3.000	3.589	6.000	7.250

complicated to be used for estimating the parameters of the distribution, but it can be shown numerically that convergence is rapid if ρ is not large. If h/ξ is an integer, n say, then it can be shown directly that the first three moments about the origin are exactly

$$\left. \begin{aligned} \mu &= \frac{1}{m} \left\{ n - \sum_{q=1}^n \sum_{r=0}^{n-q} \frac{(\rho q)^r}{r!} e^{-\rho q} \right\}, \\ \mu'_2 &= \frac{1}{m^2} \left\{ n(n+1) - 2 \sum_{q=1}^n \sum_{r=0}^{n-q+1} (n-q+1) \frac{(\rho q)^r}{r!} e^{-\rho q} \right\}, \\ \mu'_3 &= \frac{1}{m^3} \left\{ n(n+1)(n+2) - 3 \sum_{q=1}^n \sum_{r=0}^{n-q+2} (n-q+1)(n-q+2) \frac{(\rho q)^r}{r!} e^{-\rho q} \right\}. \end{aligned} \right\} \quad (21)$$

The exact and asymptotic values of the mean, variance and third moment about the mean are given in Table 1 for $h/\xi = 3$. It will be seen that the asymptotic value for the variance oscillates about the exact value. Now let $\rho = (x-1)^{-1}$, $\mu m = x^{-2}y$ and $h/\xi = x^{-1}z$ and denote σ^2/μ^2 by α_1 and μ_3/μ^3 by α_2 . Then the three equations given by the asymptotic expressions for the moments are

$$\left. \begin{aligned} y &= \frac{1}{2} + z, \\ \alpha_1 y^2 &= \frac{5}{4} - \frac{2}{3}x + z, \\ \alpha_2 y^3 &= \frac{11}{2} - 5x + \frac{3}{4}x^2 + (3-x)z. \end{aligned} \right\} \quad (22)$$

Hence estimates of ρ , m and h/ξ follow from the solution of

$$432\alpha_1^2 y^4 - 32(15\alpha_1 + 8\alpha_2) y^3 + 24(2 + 45\alpha_1) y^2 - 600y - 29 = 0. \quad (23)$$

In the course of our investigation equations (22) and (23) were used to calculate a table of values of ρ , μm and h/ξ for certain combinations of values of α_1 and α_2 . We should be glad to make these data available to any reader of this paper who wishes to go further into the application of the theory.

It would be possible to estimate the standard errors of the estimates of x , y and z and hence those of ρ , m and h/ξ . However, the formulae are too complicated for any assessment of the efficiency to be practical but, in any case, the only possible method of estimation appears to be that of moments.

This paper is published by permission of the National Coal Board.

REFERENCES

- SMITH, W. L. (1959). On the cumulants of renewal processes. *Biometrika*, **46**, 1-29.
WIDDER, D. V. (1946). *The Laplace Transform*. Princeton University Press.

A TEST OF HOMOGENEITY FOR ORDERED ALTERNATIVES

BY D. J. BARTHOLOMEW

University College of North Staffordshire

1. INTRODUCTION

A problem which occurs frequently in statistical analysis is that of testing whether a number of normal variates have the same mean. Given k such variates with mean values $m_1, m_2, \dots, m_k$ the null hypothesis which is tested may be written,

$$H_0: m_1 = m_2 = \dots = m_k.$$

The object of this paper is to give a new solution to this problem which has many applications, especially in the social sciences. The appropriate test of the hypothesis can only be determined when the alternative hypotheses, which are considered likely or important, have been specified. The most general form of alternative, which is the basis of almost all existing work, may be written,

$$H_1: m_1 \neq m_2 \neq \dots \neq m_k.$$

As examples we note that the standard analysis of variance tests fall into this category. The χ^2 -test for the equality of proportions in a $2 \times k$ contingency table is a further example if the numbers involved are large enough. In some applications it is possible to be more precise in the specification of the alternative. When this is the case it is advantageous to make use of this extra information to obtain more powerful tests. This paper deals with what have been called 'ordered' alternatives for which the rank order of the means is known. Thus instead of H_1 above, we shall consider

$$H_2: m_1 \geq m_2 \geq \dots \geq m_k.$$

At least one of the inequality signs must be a strict inequality so that H_0 is not included in H_2 . The practical importance of this alternative will be clear from the discussion below. For the simple case $k = 2$ the situation is met by using a single-tail test, but for $k > 2$ the distinction between one- and two-tail tests is lost. The present work may therefore be looked upon as a generalization of the single-tail test.

Attempts to meet the need for a test against ordered alternatives have already been made. Jonckheere (1954) discussed the one-way analysis of variance and gave a distribution-free test which is closely related to Kendall's coefficient of rank correlation. The need for a test of this kind arises in psychological work where qualitative characteristics can be ranked but not easily measured. Jonckheere mentions, as an example, an experiment to test the effect of stress on a task of manual dexterity. Here data would be obtained from groups of subjects working under high, medium, low and minimal stress, the null hypothesis being that stress has no effect on performance, and the alternative that increasing stress produces an increasing effect. Many similar examples occur in the analysis of social survey data where people are often grouped according to 'social class', 'living conditions', etc. Such factors can seldom be measured satisfactorily but they can be ranked according to their expected effect.

Armitage (1955), who discussed the problem in connexion with the analysis of $2 \times k$ contingency tables, drew his examples from the medical field. He was particularly interested

in linear changes in the proportions and gave tests which amount to converting the problem to one of regression, by allocating scores to the experimental groups.

Neither of the approaches to the problem mentioned above provides a general solution, since each was designed for a particular application. The new tests to be derived are based on the likelihood ratio and are applicable to any problem which can be formulated as a test for the equality of normally distributed means. They are closely related to the familiar χ^2 and F criteria which are used when the alternative is H_1 . Significance is easily tested using tabulated functions or the special tables of percentage points included in the paper.

2. THE TEST CRITERIA

2.1. Suppose that we have independent observations $x_1, x_2, \dots, x_k$, where x_i is normally distributed with mean m_i and standard deviation σ_i ($i = 1, 2, \dots, k$). It is important to be clear on the meaning of the suffices; thus x_1 is the observation from the population with the largest mean under H_2 but is not necessarily the largest observation. In this section it will be assumed that the σ 's are known, but in § 2.2 we shall include the case when $\sigma_i = \sigma/\sqrt{n_i}$ ($i = 1, 2, \dots, k$), the n 's being known sample numbers and σ being estimated from the data.

Our problem is to construct a test criterion which is particularly sensitive to the alternative H_2 . Since the χ^2 -test, used when the alternative is H_1 , can be derived by the likelihood ratio principle it seems reasonable to use the same approach when it is H_2 . Although it is not clear, in this case, whether the test derived has optimum properties, the method has a strong intuitive appeal and leads to a meaningful test.

The standard method of deriving likelihood ratio tests leads to the following form of test criterion

$$\bar{\chi}_k^2 = \sum_1^k a_i(x_i - \bar{x})^2 - \sum_1^k a_i(x_i - \hat{m}_i)^2, \quad (1)$$

where $a_i = 1/\sigma_i^2$, $\bar{x} = \sum_1^k a_i x_i / \sum_1^k a_i$ and $\hat{m}_1, \hat{m}_2, \dots, \hat{m}_k$ are the values of $m_1, m_2, \dots, m_k$ which maximize (1) subject to the condition $\hat{m}_1 \geq \hat{m}_2 \geq \dots \geq \hat{m}_k$. The symbol $\bar{\chi}_k^2$ is used because of the close link with the usual χ^2 -test which will become apparent in what follows. The effect of restricting the alternative to H_2 is to introduce the second term on the right. The values of $m_1, m_2, \dots, m_k$ which maximize (1) are those which minimize

$$U_k = \sum_1^k a_i(x_i - m_i)^2,$$

so the problem of determining $\bar{\chi}_k^2$ has been reduced to the following. Minimize U_k with respect to $m_1, m_2, \dots, m_k$, subject to the restriction $m_1 \geq m_2 \geq \dots \geq m_k$.

It will simplify the determination of the minimum of U_k if we use geometrical terminology. The first step is to discover whether the absolute minimum of U_k lies in the region H_2 ($m_1 \geq m_2 \geq \dots \geq m_k$). If this is so the restricted and unrestricted problems have the same solution, which, in this case, is obviously $\hat{m}_i = x_i$ for $i = 1, 2, \dots, k$. This solution lies in H_2 if $x_1 \geq x_2 \geq \dots \geq x_k$.

If this is not the case then it will be possible to find $x_j < x_{j+1}$ for some j . The absolute minimum would require $\hat{m}_j < \hat{m}_{j+1}$ which is outside H_2 . To avoid this we must introduce the restriction $m_j \geq m_{j+1}$ before minimizing U_k , as follows.

$$\begin{aligned} U_k &= \sum_{i \neq j, j+1} a_i(x_i - m_i)^2 + \{a_j(x_j - m_j)^2 + a_{j+1}(x_{j+1} - m_{j+1})^2\} \\ &= R + S, \quad \text{say.} \end{aligned}$$

The minimum of R is unaffected by the restriction and is therefore zero when $\hat{m}_i = x_i$. Consider now the minimum of S . The point $m_j = x_j$, $m_{j+1} = x_{j+1}$ falls outside the region $m_j \geq m_{j+1}$ in the (m_j, m_{j+1}) plane. The minimum of S will therefore occur on the boundary of the region, that is when $m_j = m_{j+1}$. In this case it is easily shown that

$$\hat{m}_j = \hat{m}_{j+1} = \frac{a_j x_j + a_{j+1} x_{j+1}}{a_j + a_{j+1}}.$$

It will be convenient, for what follows, to introduce the notation

$$X(p, q) = \frac{\sum_{i=p}^q a_i x_i}{\sum_{i=p}^q a_i}.$$

The minimum of U_k subject to the restriction $m_j \geq m_{j+1}$ may therefore be written

$$\begin{aligned}\hat{m}_j &= \hat{m}_{j+1} = X(j, j+1), \\ \hat{m}_i &= x_i \quad (i \neq j, j+1).\end{aligned}$$

The problem has now been reduced to one involving $(k-1)$ variates. If

$$x_1 \geq x_2 \geq \dots \geq x_{j-1} \geq X(j, j+1) \geq x_{j+2} \geq \dots \geq x_k,$$

the final solution has been reached. Otherwise the problem is the same as the original but with the number of variates reduced by one. It may therefore be treated in the same way by taking $X(j, j+1)$ as a single observation with weight $(a_j + a_{j+1})$. The rule for obtaining $\hat{m}_1, \hat{m}_2, \dots, \hat{m}_k$ may, therefore, be stated as follows.

Arrange the observations in the order predicted by the alternative, thus, $x_1, x_2, \dots, x_k$. If any consecutive pair x_j, x_{j+1} are not in the expected order (i.e. if $x_{j+1} > x_j$), form the average

$$X(j, j+1) = (a_j x_j + a_{j+1} x_{j+1}) / (a_j + a_{j+1}).$$

The series is now reduced to $(k-1)$ members of which $(k-2)$ are unchanged and one is the average of two original ones. Proceed exactly as before, treating $X(j, j+1)$ as a single observation but with weight $(a_j + a_{j+1})$. Continue in this manner until the resulting values are in the order predicted by H_2 . The following example illustrates the method.

Example

$$\begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \\ 60 & -73 & -141 & 7 \end{array}$$

$$a_1 = a_2 = a_3 = a_4.$$

We notice that $x_4 > x_3$ so we form their average,

$$X(3, 4) = \frac{1}{2}(-141 + 7) = -67.$$

Now $X(3, 4) > x_2$ so we calculate

$$\frac{1}{3}\{2X(3, 4) + x_2\} = X(2, 4) = \frac{1}{3}(-73 - 141 + 7) = -69.$$

As $x_1 > X(2, 4)$ the solution is

$$\hat{m}_1 = 60, \quad \hat{m}_2 = \hat{m}_3 = \hat{m}_4 = -69.$$

By calculating $\hat{m}_1, \hat{m}_2, \dots, \hat{m}_k$ in the way described $\bar{\chi}_k^2$ can be obtained by substitution in (1). It is easily shown, for estimates of the form given by this method that

$$\bar{\chi}_k^2 = \sum_{i=1}^k a_i (x_i - \bar{x})^2 - \sum_{i=1}^k a_i (x_i - \hat{m}_i)^2 = \sum_{i=1}^k a_i (\hat{m}_i - \bar{x})^2. \quad (2)$$

The difference between the method of calculation of the usual χ^2 and $\bar{\chi}_k^2$ is that for the latter we use the estimates $\hat{m}_1, \hat{m}_2, \dots, \hat{m}_k$ instead of the original observations $x_1, x_2, \dots, x_k$. This is illustrated by the following example.

<i>Example</i>	<i>i</i>	1	2	3	4
	x_i	-0.83	1.88	0.20	0.44
	a_i	5	10	10	5

Carrying out the procedure described above we find

$$\hat{m}_1 = \hat{m}_2 = 0.98, \quad \hat{m}_3 = \hat{m}_4 = 0.28, \quad \bar{x} = 0.63.$$

Hence $\bar{\chi}_4^2 = 5(0.98)^2 + 10(0.98)^2 + 10(0.28)^2 + 5(0.28)^2 - 30(0.63)^2 = 3.68$.

The example also indicates an alternative method of calculation. We could look upon the problem as involving two observations, $X(1, 2) = 0.98$, $X(3, 4) = 0.28$ each with weight $(5 + 10)$. $\bar{\chi}_4^2$ would then be calculated as the usual χ^2 sum of squares using $X(1, 2)$, $X(3, 4)$ and the appropriate weights. This alternative method is always available and is, in fact, the simplest method of calculation. When the averaging process is complete we have l means,

$$X(1, k_1), \quad X(k_1 + 1, k_1 + k_2), \dots, X(k - k_l + 1, k)$$

with weights

$$A(1, k_1), \quad A(k_1 + 1, k_1 + k_2), \dots, A(k - k_l + 1, k),$$

where

$$A(p, q) = \sum_p^q a_i, \quad \sum_1^l k_i = k.$$

$\bar{\chi}_k^2$ is then obtained as the sum of squares of these means. This will be referred to as the reduced form of the problem.

2.2. In this section we give the modifications necessary when σ can be estimated from the data. Consider the problem as it arises in the one-way analysis of variance by groups where we have observations x_{ij} ($i = 1, 2, \dots, k; j = 1, 2, \dots, n_i$), which are normally and independently distributed about a mean m_i with common standard deviation σ . Define

$$\bar{x}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \quad \text{and} \quad \bar{x}_{..} = \frac{\sum_1^k n_i \bar{x}_{i.}}{\sum_1^k n_i},$$

then we have to test whether $\bar{x}_{1.}, \bar{x}_{2.}, \dots, \bar{x}_{k.}$ come from populations with the same mean value. The mean $\bar{x}_{i.}$ has standard deviation $\sigma/\sqrt{n_i}$ and therefore $a_i = n_i/\sigma^2$. The most obvious modification of the test criterion is the replacement of σ^2 by the residual mean square obtained from the analysis of variance table. If this is denoted by S_e^2 the criterion would become

$$\sum_1^k n_i (\hat{m}_i - \bar{x}_{i.})^2 / S_e^2.$$

However, a slight improvement can be obtained as follows. Suppose, for example, that $\bar{x}_{i.} < \bar{x}_{i+1.}$, then following the rule for calculating the test criterion we form

$$X(i, i+1) = (n_i \bar{x}_{i.} + n_{i+1} \bar{x}_{i+1.}) / (n_i + n_{i+1}).$$

This is exactly equivalent to pooling the i th and $(i+1)$ th groups and then calculating the average of the single group so formed. The residual mean square calculated from this new table will have one more degree of freedom than that from the original table, because the 'between groups' degrees of freedom have been decreased by one with the total remaining

constant. In general, it can be shown that if the reduced form of the problem consists of l groups then the increase in the residual degrees of freedom will be $(k-l)$. The test criterion to be used is therefore defined as

$$\bar{F}_k = \sum_1^k n_i (\hat{m}_i - \bar{x}_{..})^2 / (l-1) S_e'^2, \quad (3)$$

where $S_e'^2$ is the residual mean square of the analysis of variance table obtained from the reduced form of the problem, having $(N-l)$ degrees of freedom $\left(N = \sum_1^k n_i\right)$. $\bar{F}_k$ is thus the ordinary mean square ratio calculated for the reduced problem. An example, illustrating the method of calculation, is included in § 5. It can readily be shown that $\bar{F}_k$ is the criterion produced by the likelihood ratio principle.

3. THE DISTRIBUTION OF $\bar{\chi}_k^2$ AND $\bar{F}_k$

3.1. *The distribution of $\bar{\chi}_k^2$.* To illustrate the method of approach to the distribution problem we consider first the case $k = 2$. This must, of course, reduce to the familiar one-tail test. We have

$$\begin{aligned} \bar{\chi}_2^2 &= \sum_1^2 a_i (x_i - \bar{x})^2 \quad \text{for } x_1 > x_2 \\ &= 0 \quad \text{for } x_2 \geq x_1. \end{aligned}$$

For tests of significance the probability required is

$$\Pr \{\bar{\chi}_2^2 \geq \gamma\} = \Pr \left\{ x_1 > x_2; \sum_1^2 a_i (x_i - \bar{x})^2 \geq \gamma \right\}$$

when the null hypothesis is true. Now the value of $\sum_1^2 a_i (x_i - \bar{x})^2$ does not depend on the order of x_1 and x_2 but only on their absolute difference so we may write

$$\Pr \{\bar{\chi}_2^2 \geq \gamma\} = \Pr \{x_1 > x_2\} \Pr \left\{ \sum_1^2 a_i (x_i - \bar{x})^2 \geq \gamma \right\}.$$

Further, since on the null hypothesis $\Pr \{x_1 > x_2\} = \frac{1}{2}$ whatever a_1, a_2 , we have finally

$$\Pr \{\bar{\chi}_2^2 \geq \gamma\} = \frac{1}{2} \Pr \{\chi_1^2 \geq \gamma\},$$

where χ_ν^2 denotes χ^2 with ν degrees of freedom. This will be recognized as the equivalent of the one-tail test.

For larger k the method is similar. Let $P(l, k; a_1, a_2, \dots, a_k)$ denote the probability that the reduced form of the problem consists of l means when the original observations $x_1, x_2, \dots, x_k$ had weights $a_1, a_2, \dots, a_k$. When there is no risk of confusion we shall suppress the a 's and write $P(l, k)$.

The probability integral of $\bar{\chi}_k^2$ may thus be written

$$\begin{aligned} \Pr \{\bar{\chi}_k^2 \geq \gamma\} &= \sum_{l=2}^k P(l, k) \Pr \{\bar{\chi}_k^2 \geq \gamma | l\}, \\ \Pr \{\bar{\chi}_k^2 = 0\} &= P(1, k), \end{aligned} \quad (4)$$

where $\Pr \{\bar{\chi}_k^2 \geq \gamma | l\}$ is the probability that $\bar{\chi}_k^2$ exceeds γ , given that it was calculated from l means. This is simply equal to $\Pr \{\chi_{l-1}^2 \geq \gamma\}$ as we now show.

An obvious simplification of the notation introduced at the end of § 2.1 gives

$$\bar{\chi}_k^2 = \sum_1^l A_i (X_i - \bar{X})^2,$$

where $\bar{X} = \sum_1^l A_i X_i / \sum_1^l A_i$ and $\text{var}(X_i) = A_i^{-1}$ ($i = 1, 2, \dots, l$).

Two conditions must be satisfied if $\bar{\chi}_k^2$, for given l , is to be distributed like χ^2 with $(l-1)$ degrees of freedom.

(i) Each X_i ($i = 1, 2, \dots, l$) must be normally distributed about a common mean.

(ii) The restriction $X_1 \geq X_2 \geq \dots \geq X_l$ must be irrelevant as far as the distribution of $\bar{\chi}_k^2$ is concerned. This amounts to showing that the X 's may be treated as if they were independent.

The first condition is satisfied, on the null hypothesis, since the averaging which leads to any X_i does not depend on the magnitude of the values which are averaged. They can therefore be regarded as being selected at random.

To show that the second condition is also satisfied, consider the following, algebraically equivalent, expression for $\bar{\chi}_k^2$.

$$\bar{\chi}_k^2 = \sum_{i>j} \sum A_i A_j (X_i - X_j)^2 / \sum_{i=1}^l A_i.$$

If the X 's were independent then $(X_i - X_j)^2$ would be the square of a normal variate with zero mean and $\bar{\chi}_k^2$ would be a linear function of such variates. When the X 's are ordered, $(X_i - X_j)$ is distributed like the modulus of a normal variable (since $X_i \geq X_j$ when $i > j$). However, the square of the modulus has exactly the same distribution as the square of the difference when the sign is not restricted. The distribution of $\bar{\chi}_k^2$ is therefore unaffected by the dependence among the X 's resulting from the restriction $X_1 \geq X_2 \geq \dots \geq X_l$.

(4) may therefore be re-written

$$\Pr\{\bar{\chi}^2 \geq \gamma\} = \sum_2^k P(l, k) \int_{\gamma}^{\infty} p(\chi_{l-1}^2) d\chi_{l-1}^2, \quad (5)$$

$$\Pr(\bar{\chi}^2 = 0) = P(1, k).$$

The distribution problem thus centres on the determination of the probabilities $P(l, k)$. In what follows we give complete results for $k = 3$ and $k = 4$, and a partial result for $k = 5$, using known results about the normal multivariate integral. The following notation is used

$$z_i = x_i - x_{i+1} \quad (i = 1, 2, \dots, (k-1)).$$

ρ_{ij} is the product moment correlation coefficient of z_i and z_j , which is found to be

$$\rho_{i, i+1} = - \sqrt{\left\{ \frac{a_i a_{i+2}}{(a_i + a_{i+1})(a_{i+1} + a_{i+2})} \right\}} \quad \text{for } i = 1, 2, \dots, (k-1),$$

$$\rho_{ij} = 1 \quad \text{for } i = j,$$

$$= 0 \quad \text{for } |i-j| > 1. \quad (6)$$

$k = 3$

$$P(3, 3) = \Pr\{z_1 > 0, z_2 > 0\}.$$

On the null hypothesis z_1 and z_2 have a normal bivariate distribution with zero means and correlation coefficient ρ_{12} ; hence using the well-known result of Sheppard

$$P(3, 3) = \frac{1}{2} \left[1 - \frac{\cos^{-1} \rho_{12}}{\pi} \right].$$

In the special case $a_1 = a_2 = a_3, \rho_{12} = -\frac{1}{2}$ and $P(3, 3) = 1/3!$; a result which could have been obtained directly since, in this case, all permutations of the x 's are equally likely. $P(2, 3)$ can be obtained using the known results for $k = 2$. Two possibilities must be considered:

$$(i) \quad x_1 < x_2, \quad X(1, 2) > x_3$$

$$(ii) \quad x_2 < x_3, \quad X(2, 3) < x_1.$$

Now $X(1, 2)$ is normally distributed with zero mean and variance $(a_1 + a_2)^{-1}$, so that the event $x_1 < x_2$ is independent of the event $X(1, 2) > x_3$. Hence the first of these possibilities has probability

$$P(1, 2; a_1, a_2) P(2, 2; (a_1 + a_2), a_3).$$

By a similar argument the probability of the second is

$$P(1, 2; a_2, a_3) P(2, 2; a_1, (a_2 + a_3)).$$

Because $P(1, 2) = P(2, 2) = \frac{1}{2}$ whatever the weights,

$$P(2, 3) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Using the fact that $\sum_{l=1}^3 P(l, k) = 1$, $P(1, 3)$ is found to be

$$\frac{1}{2\pi} \cos^{-1} \rho_{12}.$$

$k = 4$

$$P(4, 4) = \Pr\{z_1 > 0, z_2 > 0, z_3 > 0\}.$$

When z_1, z_2, z_3 are distributed in the trivariate normal form with zero means, David (1953) gives the result

$$P(4, 4) = \frac{1}{4\pi} (2\pi - \cos^{-1} \rho_{12} - \cos^{-1} \rho_{23} - \cos^{-1} \rho_{13})$$

the ρ_{ij} 's being given by (6).

$P(3, 4)$ is built up, by a similar argument to that used for $k = 3$, using the results for $k = 2$ and 3 as follows:

$$P(3, 4) = P(1, 2; a_1, a_2) P(3, 3; (a_1 + a_2), a_3, a_4) + P(1, 2; a_2, a_3) P(3, 3; a_1, (a_2 + a_3), a_4) \\ + P(1, 2; a_3, a_4) P(3, 3; a_1, a_2, (a_3 + a_4)).$$

$$\text{Therefore} \quad P(3, 4) = \frac{1}{4} \left[\left(1 - \frac{\cos^{-1} \rho'_{12}}{\pi}\right) + \left(1 - \frac{\cos^{-1} \rho''_{12}}{\pi}\right) + \left(1 - \frac{\cos^{-1} \rho'''_{12}}{\pi}\right) \right].$$

The correlations, which are given below, prove to be equal to the partial correlations between the z 's, as indicated.

$$\rho'_{12} = - \sqrt{\frac{(a_1 + a_2) a_4}{(a_1 + a_2 + a_3)(a_3 + a_4)}} = \rho_{23 \cdot 1}$$

$$\rho''_{12} = - \sqrt{\frac{a_1 a_4}{(a_1 + a_2 + a_3)(a_2 + a_3 + a_4)}} = \rho_{13 \cdot 2}$$

$$\rho'''_{12} = - \sqrt{\frac{a_1(a_3 + a_4)}{(a_1 + a_2)(a_2 + a_3 + a_4)}} = \rho_{12 \cdot 3},$$

whence

$$P(3, 4) = \frac{3}{4} - (\cos^{-1} \rho_{12 \cdot 3} + \cos^{-1} \rho_{23 \cdot 1} + \cos^{-1} \rho_{13 \cdot 2}) / 4\pi.$$

In the same manner the equation for $P(2, 4)$ is found to be

$$P(2, 4) = P(1, 3; a_1, a_2, a_3) P(2, 2; (a_1 + a_2 + a_3), a_4) + P(1, 3; a_2, a_3, a_4) P(2, 2; a_1, (a_2 + a_3 + a_4)) \\ + P(1, 2; a_1, a_2) P(1, 2; a_3, a_4) P(2, 2; (a_1 + a_2), (a_3 + a_4)).$$

Substituting known results in this equation we find

$$P(2, 4) = \frac{1}{8} + [\cos^{-1} \rho_{12} + \cos^{-1} \rho_{23}] / 4\pi \\ = \frac{1}{2} - P(4, 4).$$

Finally, by subtraction

$$P(1, 4) = (\cos^{-1} \rho_{12.3} + \cos^{-1} \rho_{23.1} + \cos^{-1} \rho_{13.2}) / 4\pi - \frac{1}{4}.$$

$k > 4$

It is not, in general, possible to obtain the distribution for $k > 4$ because there is no known expression for $\Pr\{z_1 > 0, z_2 > 0, \dots, z_m > 0\}$ when $m > 3$. Some numerical values and methods have, however, been given by Plackett (1954) and these may prove useful. For the special, but important, case when the α 's are all equal, results can be obtained for $k = 5$. In this case all permutations of the x 's are equally probable and so

$$\Pr(x_1 > x_2 > x_3 > x_4 > x_5) = P(5, 5) = 1/5! = \frac{1}{120}.$$

The probabilities required to obtain $P(l, 5)$ for $l < 5$ are all known so the above methods can be used. The results obtained are given in Table 1. I have conjectured the following result for the general case with equal weights*. It yields the correct values for $k \leq 5$.

$$P(l, k) = \sum_{j=l-1}^{k-1} P(1, k-j) P(l-1, j) / l.$$

The proof is straightforward for $l = 2$ and $l = k$, but breaks down for the intermediate cases.

Table 1. Values of $P(l, k)$ when the weights are equal

l	$2P(l, 2)$	$6P(l, 3)$	$24P(l, 4)$	$120P(l, 5)$
1	1	2	6	24
2	1	3	11	50
3	—	1	6	35
4	—	—	1	10
5	—	—	—	1

3.2. *The distribution of $\bar{F}_k$.* The conditional distribution of $\bar{F}_k$, given that there are l means in the reduced form of the problem, is simply that of $\bar{F}_{l-1, N-l}$ by the argument of § 3.1. The probabilities $P(l, k)$ depend only on the correlation coefficients ρ_{ij} which are independent of σ^2 . Therefore

$$\Pr\{\bar{F}_k \geq \gamma\} = \sum_2^k P(l, k) \Pr\{\bar{F}_{l-1, N-l} \geq \gamma\}, \quad (7)$$

$$\Pr\{\bar{F}_k = 0\} = P(1, k).$$

For testing significance it is convenient to transform the F distribution to an incomplete B -function giving

$$\Pr\{\bar{F}_k \geq \gamma\} = \sum_2^k P(l, k) (1 - I_z\{\frac{1}{2}(l-1), \frac{1}{2}(N-l)\}) = \sum_2^k P(l, k) I_{1-z}\{\frac{1}{2}(N-l), \frac{1}{2}(l-1)\}, \quad (8)$$

where

$$z = (l-1)\gamma / (N-l + (l-1)\gamma).$$

* See Note added in proof on p. 48.

4. CALCULATION OF PERCENTAGE POINTS

The results of § 3 give all that is necessary for the calculation of percentage points for the $\bar{\chi}_k^2$ and $\bar{F}_k$ statistics. Given the weights $a_1, a_2, \dots, a_k$, and tables of the χ^2 integral or incomplete B -function, it is simply a matter of substitution in (5) or (8). The purpose of this section is to give tables of percentage points to reduce the work involved in testing significance.

Consider first the distribution of $\bar{\chi}_3^2$. From (5) we have

$$\Pr\{\bar{\chi}_3^2 \geq \gamma\} = \frac{1}{2} \left(1 - \frac{\cos^{-1} \rho_{12}}{\pi}\right) \int_{\gamma}^{\infty} p(\chi_2^2) d\chi_2^2 + \frac{1}{2} \int_{\gamma}^{\infty} p(\chi_1^2) d\chi_1^2.$$

Table 2 gives the values of γ satisfying

$$\begin{aligned} \Pr\{\bar{\chi}_3^2 \geq \gamma\} &= 0.05 \\ &= 0.01, \quad \text{for } -\rho_{12} = 0(0.1) 1.0. \end{aligned}$$

This table will be adequate for most practical work, using rough interpolation if necessary. When high accuracy is required it is better to make a direct calculation of the probability rather than use interpolation methods. This remark applies also to Table 3.

Table 2. The 5 and 1 % points of $\bar{\chi}_3^2$

$-\rho_{12}$	5 %	1 %	$-\rho_{12}$	5 %	1 %
0.0	4.231	7.289	0.6	3.715	6.700
0.1	4.158	7.208	0.7	3.593	6.556
0.2	4.081	7.122	0.8	3.446	6.377
0.3	4.001	7.030	0.9	3.245	6.130
0.4	3.914	6.932	1.0	2.706	5.413
0.5	3.820	6.822	—	—	—

Table 3 gives the 5 and 1 % points of $\bar{\chi}_4^2$ for $-\rho_{12}, -\rho_{23} = 0(0.1) 1.0$, subject to the requirement

$$-\sqrt{(1-\rho_{23}^2)} < \rho_{12} < \sqrt{(1-\rho_{23}^2)}.$$

This restriction is simply a consequence of the fact that partial correlations must lie in the range $(-1, +1)$. As the probability integral is symmetrical in ρ_{12}, ρ_{23} the lower half only of the table is printed. Values in the upper half are obtained by interchanging ρ_{12} and ρ_{23} .

The 5 and 1 % points for equal weights when $k = 5$ are given in Table 4.

In each case the probability integral was calculated from the distribution function of χ^2 as tabulated in *Biometrika Tables for Statisticians*, and from tables of $\sin^{-1} x$. The percentage points were then obtained by inverse interpolation using the second order Bessel formula. They were checked by direct calculation using exact expressions for the χ^2 integral.

Percentage points for the $\bar{F}_k$ distribution have not yet been calculated. Although feasible for $k = 3$ a table with 3-way entry would be required for $k = 4$ and one with 4-way entry for $k = 5$. It is planned to carry out further work on the distribution of $\bar{F}_k$ to obtain, if possible, simple approximations. This would avoid the necessity for extensive tabulations.

Table 3. *The 5 % (upper figure) and 1 % (lower figure) points of $\bar{\chi}_4^2$*

$-\rho_{12} \backslash -\rho_{23}$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.0	5.435 8.746	— —	— —	— —	— —	— —	— —	— —
0.1	5.372 8.676	5.305 8.600	— —	— —	— —	— —	— —	— —
0.2	5.307 8.602	5.235 8.521	5.160 8.436	— —	— —	— —	— —	— —
0.3	5.239 8.525	5.162 8.438	5.081 8.346	4.993 8.247	— —	— —	— —	— —
0.4	5.167 8.442	5.083 8.348	4.994 8.247	4.898 8.137	4.791 8.015	— —	— —	— —
0.5	5.088 8.351	4.997 8.248	4.898 8.137	4.791 8.014	4.670 7.873	4.528 7.709	— —	— —
0.6	5.003 8.250	4.900 8.135	4.789 8.008	4.665 7.867	4.524 7.703	4.354 7.505	4.135 7.244	— —
0.7	4.904 8.135	4.787 8.002	4.657 7.854	4.510 7.684	4.337 7.482	4.117 7.224	3.805 6.846	3.137 6.001
0.8	4.787 7.994	4.644 7.831	4.484 7.647	4.294 7.427	4.056 7.147	3.715 6.734	2.706 5.413	— —
0.9	4.631 7.804	4.432 7.579	4.196 7.303	3.883 6.934	3.353 6.278	— —	— —	— —
1.0	4.231 7.289	— —	— —	— —	— —	— —	— —	— —

Table 4. *The 5 and 1 % points of $\bar{\chi}_5^2$ for equal weights*

5 %	1 %
5.049	8.356

5. APPLICATION OF THE TESTS

In this section three sets of data are given to which the $\bar{\chi}_k^2$ or $\bar{F}_k$ tests can be applied. The first two relate to $2 \times k$ contingency tables and are taken from the paper by Armitage (1955); the third is an artificial analysis of variance table. The justification for applying the test to $2 \times k$ contingency tables is as follows. Suppose that out of n_i observations in the i th category a proportion p_i show a certain characteristic; it is desired to test whether these proportions are equal. If each n_i is large and p_i not too close to 0 or 1 we can assume that p_i is normally distributed with variance $P(1-P)/n_i$, where P is the proportion showing the characteristic in the whole sample. The problem is thus in a form suitable for the application of the $\bar{\chi}_k^2$ -test

with $a_i = n_i / \{P(1 - P)\}$ ($i = 1, 2, \dots, k$). This approximation is, of course, the basis of the χ^2 -test and is usually considered to be adequate if no expected cell frequency is less than, say, 5.

Example 1

The data shown in Table 5 are given in Armitage's paper, where references to the original work may be found. The null hypothesis is that the chance of being a carrier does not depend

Table 5. Relationship between nasal carrier rate for streptococcus pyogenes and size of tonsils among 1398 children aged 0-15 years

	Present but not enlarged +	Enlarged tonsils		Total
		++	+++	
Carriers	19	29	24	72
Non-carriers	497	560	269	1326
Total	516	589	293	1398
Carrier rate	0.0368	0.0492	0.0819	—

Table 6. Changes in the size of ulcer crater, 3 months after treatment, for patients in two treatment groups (A and B)

Treatment group	Number of cases with crater				Total
	Larger	$< \frac{2}{3}$ healed	$\geq \frac{2}{3}$ healed	Healed	
A	6	4	10	12	32
B	11	8	8	5	32
n_i	17	12	18	17	64
Proportion in group A	0.353	0.333	0.556	0.706	—

on size of tonsils; the alternative that the chance increases with size of tonsils. We note that the proportions are in the order expected under the alternative and hence $\bar{\chi}_3^2$ has the same value as χ^2 for this table.

Thus $\bar{\chi}_3^2 = 7.863$,

$$-\rho_{12} = \sqrt{\left\{ \frac{516 \times 293}{(516 + 589)(589 + 293)} \right\}} = 0.394.$$

Referring to Table 2 this is found to be significant at the 1% point in contrast to the customary χ^2 , for which the significance lies between $2\frac{1}{2}$ and 1%.

Example 2

Table 6, also from Armitage's paper, gives the results of an experiment to compare two methods of ulcer treatment.

We test the null hypothesis that the two treatments are equally good, against the alternative that treatment A is better than treatment B. Under this alternative we should expect

the proportions to increase as we move from left to right across the table. In this example the proportions are not in the expected order to begin with. Because the first is greater than the second, the first two groups are combined into one giving a table with three groups. This new table has the proportions in the expected order. $\bar{\chi}_4^2$ is obtained by calculating χ^2 for this pooled table, giving $\bar{\chi}_4^2 = 5.900$, $-\rho_{12} = 0.593$, $-\rho_{23} = 0.441$.

Extracting the relevant part of Table 3, remembering that ρ_{12} and ρ_{23} can be interchanged, we have for the 5 % points,

$-\rho_{23} \backslash -\rho_{12}$	0.5	0.6
	4.670	4.524
0.4	4.528	4.354
0.5		

$\bar{\chi}_4^2$ is therefore clearly significant at the 5 % level (but not at 1 %). If the χ^2 -test had been used the significance would have been between 11 and 12 %.

Example 3

The following table gives artificial data constructed to demonstrate the application of the $\bar{F}_k$ -test to an analysis of variance table.

Table 7. Artificial data to illustrate the application of the $\bar{F}_k$ test

Group ...	I	II	III	IV	V	Total
Total	110	123	126	140	96	595
No. observed	4	4	4	4	4	20
Sum of squares	3,380	3,601	3,712	3,984	3,201	12,378
Mean	27.50	30.75	31.50	35.00	24.00	—
				29.50		
				30.17		
			30.31			

It is supposed that the alternative predicts that the means will increase as we move from left to right across the table. We begin by averaging the means of groups IV and V followed by III, IV and V and finally II, III, IV and V as indicated underneath the table. The reduced form of the problem therefore consists of two groups; group I with mean 27.50 and groups II-V with mean 30.31. Calculating the F ratio for the reduced problem we obtain

$$F = 3.008,$$

which is the value of $\bar{F}_k$ for Table 7. To test the significance we require, from (8)

$$z = \frac{3.008}{3.008 + 18} = 0.1432.$$

The following table sets out the method of calculating $\Pr\{\bar{F}_5 \geq 3.008\}$. As the numbers in each group are equal the probabilities $P(l, k)$ are obtained directly from Table 1.

There is thus no reason to reject the null hypothesis that the groups are drawn from populations with the same mean.

Table 8. Calculations required for testing the significance of $\bar{F}_k$

l	$P(l, k)$	$I_{1-\alpha}(\frac{1}{2}(l-1), \frac{1}{2}(N-l))$	
		$z = 0.14$	$z = 0.15$
2	50/120	0.1041	0.0916
3	35/120	0.2775	0.2512
4	10/120	0.4779	0.4439
5	1/120	0.6614	0.6281

$$\Pr(z \geq 0.14) = 0.1696, \quad \Pr(z \geq 0.15) = 0.1537.$$

6. FURTHER WORK

As pointed out at the end of § 4 the greatest need is for tables and approximate methods to simplify the method for testing the significance of $\bar{F}_k$. It would also be an advantage to obtain values of $P(l, k)$ for $k \geq 5$. If the conjecture given in § 3 above can be proved, then the distribution will be completely determined for the case of equal weights. Until this problem has been solved the best course appears to be the reduction of the number of groups by pooling. This will undoubtedly result in a loss of efficiency but even so may be better than the usual χ^2 -test, which takes no account of the order. The answer to this question would require the calculation of the power function of $\bar{\chi}_k^2$. A new problem arises in the analysis of data like that of example 2 in § 4. When applying the $\bar{\chi}_k^2$ -test it was assumed that treatment A would be the better of the two treatments. If we had wished to test for a difference without specifying which would be the better the alternative would have had to be,

$$H_3: m_1 \geq m_2 \geq \dots \geq m_k,$$

$$\text{or } m_1 \leq m_2 \leq \dots \leq m_k.$$

It is hoped to communicate the results of work on these topics later.

I am indebted to the referee for suggesting a shorter method of deriving the test criterion and for pointing out a gap in the original derivation of equation (5).

Editorial Note added in proof. The recurrence relation for the probabilities $P(l, k)$ conjectured in § 3.1 has now been proved by Mr R. E. Miles. He has also established the more useful recurrence relation

$$P(l, k) = [P(l-1, k-1) + (k-1)P(l, k-1)]/k$$

for $2 \leq l \leq k-1$, which, together with $P(1, k) = k^{-1}$ and $P(k, k) = (k!)^{-1}$, determine the probabilities easily. It is hoped to publish the proof in the next issue of *Biometrika*. This result can now be used for testing the significance of $\bar{\chi}_k^2$ and $\bar{F}_k$ for $k > 5$ if the weights are equal.

REFERENCES

- ARMITAGE, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics*, **11**, 375-86.
 DAVID, F. N. (1953). A note on the evaluation of the multivariate normal integral. *Biometrika*, **40**, 458-9.
 JONCKHEERE, A. R. (1954). A distribution-free k -sample test against ordered alternatives. *Biometrika*, **41**, 133-45.
 PLACKETT, R. L. (1954). A reduction formula for normal multivariate integrals. *Biometrika*, **41**, 351-60.

SOME PROBLEMS INVOLVING LINEAR HYPOTHESES IN MULTIVARIATE ANALYSIS

BY C. RADHAKRISHNA RAO

Indian Statistical Institute, Calcutta

1. INTRODUCTION

We consider p correlated normal variables $y_1, \dots, y_p$ with the following specification.

$$(i) \quad E(y_i) = a_{i1}\tau_1 + \dots + a_{im}\tau_m, \quad (1.1)$$

or in matrix notation $E(\mathbf{y}) = \mathbf{A}\boldsymbol{\tau}$, where $\mathbf{y}$ is the column vector of the random variables $y_1, \dots, y_p$; $\boldsymbol{\tau}$ that of m unknown parameters $\tau_1, \dots, \tau_m$ and $\mathbf{A}$, the matrix of coefficients is known.

(ii) The dispersion matrix of $y_1, \dots, y_p$ is $e^{-1}\mathbf{\Lambda}$ where e is a known constant and $\mathbf{\Lambda}$ is unknown, but an estimate $\mathbf{S}$ of $\mathbf{\Lambda}$ is available and has Wishart's distribution based on f degrees of freedom independently of $y_1, \dots, y_p$. The matrix $\mathbf{\Lambda}$ is assumed to be non-singular.

The set-up is thus seen to be that of the theory of least squares with correlated variables generalized to the case where the variances and covariances are unknown, but an estimate is available. The estimates of parameters as well as the test criteria suggested in this paper are essentially those derived by the theory of least squares, substituting the estimate $\mathbf{S}$ for the unknown $\mathbf{\Lambda}$. Fortunately, the distribution problems present no difficulty and valid inferences can be drawn with the help of the existing t and F tables only.

A simple case where such a set-up holds is when we have a sample of size n from a p -variate normal distribution with an unknown dispersion matrix $\mathbf{\Lambda}$ and the mean of x_i , the i th variable, is as specified in (1.1). Here

$$y_i = \bar{x}_i, \quad e = \frac{1}{n} \quad (1.2)$$

and $\mathbf{S}$ is the matrix of estimated variances and covariances on $f = (n-1)$ degrees of freedom.

In the general set-up, on the basis of a given realization of $y_1, \dots, y_p$ and $\mathbf{S}$ we wish to examine the following problems.

(i) Is the specification (1.1) adequate? Thus, if $y_1, \dots, y_p$ represent the average weights of a group of individuals at p time points, we may ask whether the average growth curve can be represented by a polynomial in time of a given degree. In such a case the parameters represent the coefficients of the polynomial trend.

(ii) If (1.1) is true, how can estimates of $\tau_1, \dots, \tau_m$ be obtained and the precision of estimates expressed?

(iii) Assuming the specification (1.1), how can general linear hypotheses concerning $\tau_1, \dots, \tau_m$ be tested?

(iv) How can simultaneous confidence limits to a class of linear functions of $\tau_1, \dots, \tau_m$ be obtained?

Problem (i) is a generalization of some particular cases considered earlier, such as testing the equality of means of p -correlated variables, equality of means of subsets of variables, etc. (Hsu, 1938; Rao, 1946*b*, 1952). Problem (iv), which is the main motivation for the present

paper, is an attempt to obtain confidence bands of a smaller width (see § 4) than those given by Hoel (1954) in the case of linear regression such as polynomial growth curves, dosage-response relationships, etc., which are covered by the general set-up (1.1).

2. TEST FOR ADEQUACY OF THE SPECIFICATION (1.1)

Let $\mathbf{B}$ be a matrix with r , equal to p minus the rank of $\mathbf{A}$, independent vectors

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{21} & \dots & b_{p1} \\ \cdot & \cdot & \dots & \cdot \\ b_{1r} & b_{2r} & \dots & b_{pr} \end{pmatrix} \quad (2.1)$$

such that the vectors in $\mathbf{B}$ are orthogonal to those in $\mathbf{A}'$. The computation of $\mathbf{B}$, known as the deficiency matrix of $\mathbf{A}'$, can be conveniently carried out by the method of sweep out (Rao, 1952, § 1a.3). It is easy to see that the variables $z_1, \dots, z_r$ defined by

$$\left. \begin{aligned} z_1 &= b_{11}y_1 + \dots + b_{p1}y_p \\ \cdot & \quad \cdot \quad \dots \quad \cdot \\ z_r &= b_{1r}y_1 + \dots + b_{pr}y_p \end{aligned} \right\} \quad (2.2)$$

are such that

$$E(z_i) = 0, \quad (i = 1, \dots, r) \quad (2.3)$$

independently of the unknown parameters τ , and with dispersion matrix equal to $e^{-1}\mathbf{B}\mathbf{A}\mathbf{B}'$. Given $y_1, \dots, y_p$ and $\mathbf{S}$ we can compute $z_1, \dots, z_r$ by (2.2) and their estimated dispersion matrix $e^{-1}\mathbf{B}\mathbf{S}\mathbf{B}' = e^{-1}\mathbf{G}$ (say). It is easy to see that $\mathbf{G}$ itself has a Wishart distribution based on the same degrees of freedom f as that of $\mathbf{S}$ and is distributed independently of $z_1, \dots, z_r$.

The problem of testing the adequacy of the specification (1.1) is the same as that of testing whether the r correlated variables defined in (2.2) have zero mean values. The appropriate statistic for this is Hotelling's T defined by

$$T_r = \frac{e}{f} \mathbf{z}' \mathbf{G}^{-1} \mathbf{z} = \frac{e}{f} \sum_{ij} g^{ij} z_i z_j \quad (2.4)$$

with the corresponding variance ratio (Rao, 1952, p. 71)

$$\frac{f+1-r}{r} T_r \quad (2.5)$$

based on r and $(f+1-r)$ degrees of freedom. The matrix $\mathbf{B}$ is not unique but the statistic T_r is invariant as we shall see below.

An alternative way of deriving the test criterion is to consider first the statistic appropriate to test the specification (1.1) when $\mathbf{\Lambda}$ is known. This according to the least squares theory is the residual quadratic form

$$\min e(\mathbf{y} - \mathbf{A}\tau)' \mathbf{\Lambda}^{-1}(\mathbf{y} - \mathbf{A}\tau) \quad (2.6)$$

obtained by fitting the parameters by the method of least squares. This residual quadratic form (2.6) for known $\mathbf{\Lambda}$ is distributed as χ^2 on r degrees of freedom. When $\mathbf{\Lambda}$ is unknown we may consider the statistic

$$\min e(\mathbf{y} - \mathbf{A}\tau)' \mathbf{S}^{-1}(\mathbf{y} - \mathbf{A}\tau) \quad (2.7)$$

obtained by substituting the estimate $\mathbf{S}$ for $\mathbf{\Lambda}$ in (2.6).

Given the matrix $\mathbf{B}$ of (2.1) it is easy to see that there exists a matrix $\mathbf{F}$ of rank $(p-r)$ and order $(p-r, r)$ such that $\mathbf{B}\mathbf{S}\mathbf{F}' = 0$, and with such a choice of $\mathbf{F}$ (Rao, 1952, p. 53)

$$(\mathbf{y} - \mathbf{A}\tau)' \mathbf{S}^{-1}(\mathbf{y} - \mathbf{A}\tau) = (\mathbf{B}\mathbf{y})' (\mathbf{B}\mathbf{S}\mathbf{B}')^{-1}(\mathbf{B}\mathbf{y}) + (\mathbf{F}\mathbf{y} - \mathbf{F}\mathbf{A}\tau)' (\mathbf{F}\mathbf{S}\mathbf{F}')^{-1}(\mathbf{F}\mathbf{y} - \mathbf{F}\mathbf{A}\tau) \quad (2.8)$$

which is the sum of two positive definite quadratic forms with the first expression independent of τ . By choosing τ such that $Fy = FA\tau$, the second expression becomes zero, leading to the result

$$\min (y - A\tau)' S^{-1}(y - A\tau) = (By)' (BSB')^{-1} (By) = z' G^{-1} z. \quad (2.9)$$

Thus the statistic (2.7) is same as T_r defined in (2.4) apart from a constant multiplier. Incidentally the invariance of T_r with respect to the choice of B is also established.

We, thus, have two methods of computing the statistic T_r either to find a B and compute the quadratic form $z' G^{-1} z$ or obtain the minimum value in (2.6) directly by some other method.

3. TESTS OF LINEAR HYPOTHESES

In this section we consider the problem of testing linear hypotheses assuming the specification (1.1) to be true. Let

$$\left. \begin{aligned} u_{11}\tau_1 + \dots + u_{1m}\tau_m &= \xi_1, \\ &\vdots \\ u_{k1}\tau_1 + \dots + u_{km}\tau_m &= \xi_k \end{aligned} \right\} \quad (3.1)$$

or in matrix notation $U\tau = \xi$, where U and ξ are known, be a set of linear hypotheses to be tested. The rank of U is taken to be k . Let us suppose that there exist linear functions $z_{r+1}, \dots, z_{r+k}$ of $y_1, \dots, y_p$ such that

$$E(z_{r+i}) = u_{i1}\tau_1 + \dots + u_{im}\tau_m \quad (i = 1, \dots, k) \quad (3.2)$$

A linear function $u_1\tau_1 + \dots + u_m\tau_m$ is said to be not estimable if it cannot be expressed as an expectation of a linear function of the variables $y_1, \dots, y_p$. A linear hypothesis cannot be expected when the parametric function associated with it is not estimable, since such a parametric function can assume any arbitrary value without altering the probability of the observations. In the case of a set of linear hypotheses such as (3.1) each of the associated linear functions must be estimable (Rao, 1952, p. 83), the condition for which is that there exists a matrix L of order (k, p) such that $U = LA$, where A is the matrix of (1.1).

The hypothesis (3.1) together with the specification (1.1) is equivalent to the hypothesis

$$E(z_{r+1}) = \xi_1, \dots, E(z_{r+k}) = \xi_k \quad (3.3)$$

given that

$$E(z_1) = 0, \dots, E(z_r) = 0, \quad (3.4)$$

where $z_1, \dots, z_r$ are as defined in (2.2). Independent estimates of the variances and covariances of $z_1, \dots, z_{r+k}$ derivable from S have a Wishart distribution on the same degrees of freedom. The problem of testing a hypothesis such as (3.3) given (3.4) was discussed in earlier papers (Rao, 1946b, 1949). The appropriate test is as follows.

Let D be the transformation matrix defining $z_1, \dots, z_{r+k}$ in terms of $y_1, \dots, y_p$. The estimated variance-covariance matrix of $z_1, \dots, z_{r+k}$ is $DSD' = H$ (say). If $H^{-1} = (h^{ij})$ is the reciprocal matrix, then the statistic for testing given values of expectations of all the $(r+k)$ variables is

$$T_{r+k} = \frac{e}{f} \sum_{i,j=1}^{r+k} h^{ij} [z_i - E(z_i)] [z_j - E(z_j)]. \quad (3.5)$$

The statistic T_r for testing the expectations (3.4) is already defined in (2.4). The statistic for testing the hypothesis (3.3) given (3.4) is

$$\frac{f+1-k-r}{k} \frac{T_{r+k} - T_r}{1 + T_r}, \quad (3.6)$$

which is a variance ratio based on k and $(f+1-k-r)$ degrees of freedom (Rao, 1952, p. 83). We observe that the value of (3.6) would be the same for all choices of linear functions $z_{r+1}, \dots, z_{r+k}$ such that their expectations are as in (3.2). The optimum properties of a test of the type (3.6) is discussed in an earlier paper (Rao, 1949).

An alternative method of computing T_{r+s} , without necessarily making a transformation to new variables, is provided by the following equation

$$\frac{f}{e} T_{r+k} = \min (\mathbf{y} - \mathbf{A}\boldsymbol{\tau})' \mathbf{S}^{-1} (\mathbf{y} - \mathbf{A}\boldsymbol{\tau}), \quad (3.7)$$

where the minimization is subject to conditions (3.1) on $\boldsymbol{\tau}$; whereas

$$\frac{f}{e} T_r = \min (\mathbf{y} - \mathbf{A}\boldsymbol{\tau})' \mathbf{S}^{-1} (\mathbf{y} - \mathbf{A}\boldsymbol{\tau}), \quad (3.8)$$

the minimization being unrestricted.

A third method of computing T_{r+k} or directly the difference $T_{r+k} - T_r$ is provided by the theory of least squares (Rao, 1946a). This makes use of the estimates of $\boldsymbol{\tau}$ obtained by minimizing, unconditionally, $(\mathbf{y} - \mathbf{A}\boldsymbol{\tau})' \mathbf{S}^{-1} (\mathbf{y} - \mathbf{A}\boldsymbol{\tau})$ for which the normal equations are

$$\mathbf{A}' \mathbf{S}^{-1} \mathbf{y} = \mathbf{A}' \mathbf{S}^{-1} \mathbf{A} \hat{\boldsymbol{\tau}}, \quad (3.9)$$

where $\hat{\boldsymbol{\tau}}$ represents the estimates. Let $\mathbf{C}$ denote the reciprocal of $\mathbf{A}' \mathbf{S}^{-1} \mathbf{A}$, if it exists, or a pseudo-inverse as defined by Rao (1955), and

$$d_i = u_{i1} \hat{\tau}_1 + \dots + u_{im} \hat{\tau}_m \quad (i = 1, \dots, k). \quad (3.10)$$

Then

$$\frac{f}{e} (T_{r+k} - T_r) = (\mathbf{d} - \boldsymbol{\xi})' (\mathbf{U} \mathbf{C} \mathbf{U}')^{-1} (\mathbf{d} - \boldsymbol{\xi}) \quad (3.11)$$

where $\mathbf{U}$ and $\boldsymbol{\xi}$ are given in (3.1). Substituting for the numerator in (3.6) the right-hand side expression of (3.11) we find that

$$\frac{(f+1-k-r)e}{kf} \frac{(\mathbf{d} - \boldsymbol{\xi})' (\mathbf{U} \mathbf{C} \mathbf{U}')^{-1} (\mathbf{d} - \boldsymbol{\xi})}{1 + T_r} \quad (3.12)$$

has a variance ratio distribution on k and $(f+1-k-r)$ degrees of freedom.

4. SIMULTANEOUS CONFIDENCE LIMITS TO A CLASS OF LINEAR PARAMETRIC FUNCTIONS

Let us first consider the problem of setting up confidence limits to a single estimable parametric function

$$\lambda = u_1 \tau_1 + \dots + u_m \tau_m. \quad (4.1)$$

Using the notation adopted in §2, we find from (3.12) that if $d = u_1 \hat{\tau}_1 + \dots + u_m \hat{\tau}_m$ and $g = \mathbf{u}' \mathbf{C} \mathbf{u}$

$$(d - \lambda) \sqrt{\left(\frac{e(f-r)}{gf(1+T_r)} \right)} \quad (4.2)$$

has a t -distribution on $(f-r)$ degrees of freedom. Hence, the confidence limits to λ are

$$d \pm t \left(\frac{(f-r)e}{gf(1+T_r)} \right)^{-\frac{1}{2}}. \quad (4.3)$$

In particular, confidence limits to any single parameter τ_i can be obtained by choosing $\lambda = \tau_i$ provided it is estimable. The confidence limits to τ_i are

$$\hat{\tau}_i \pm t \left(\frac{e(f-r)}{c_{ii}f(1+T_r)} \right)^{-\frac{1}{2}}, \quad (4.4)$$

where c_{ii} is the i th diagonal element in the matrix $\mathbf{C}$.

Now we let the vector $\mathbf{u}$ belong to the k ($\leq m$) dimensional subspace spanned by the vectors

$$\mathbf{U} = \begin{pmatrix} u_{11} & \dots & u_{1m} \\ \vdots & \dots & \vdots \\ u_{k1} & \dots & u_{km} \end{pmatrix} \quad (4.5)$$

and represent, as in (3.10),

$$d_i - \lambda_i = (u_{i1}\hat{\tau}_1 + \dots + u_{im}\hat{\tau}_m) - \lambda_i. \quad (i = 1, \dots, k). \quad (4.6)$$

Then it is known that (Rao, 1946a, 1952, p. 85),

$$\max_{\mathbf{u}} \frac{(\mathbf{u}'\hat{\boldsymbol{\tau}} - \mathbf{u}'\boldsymbol{\tau})^2}{\mathbf{u}'\mathbf{C}\mathbf{u}} = (\mathbf{d} - \boldsymbol{\lambda})' (\mathbf{U}\mathbf{C}\mathbf{U}')^{-1} (\mathbf{d} - \boldsymbol{\lambda}) \quad (4.7)$$

when $\mathbf{u}$ is restricted to the subspace (4.5). Using the distribution result (3.12)

$$\max_{\mathbf{u}} \frac{(\mathbf{u}'\hat{\boldsymbol{\tau}} - \mathbf{u}'\boldsymbol{\tau})^2}{\mathbf{u}'\mathbf{C}\mathbf{u}} \frac{(f+1-k-r)e}{fk(1+T_r)} \quad (4.8)$$

has an F -distribution with k and $(f+1-k-r)$ degrees of freedom. If F_α represents the α -value of F corresponding to these degrees of freedom then the probability that

$$\frac{(\mathbf{u}'\hat{\boldsymbol{\tau}} - \mathbf{u}'\boldsymbol{\tau})^2}{\mathbf{u}'\mathbf{C}\mathbf{u}} \leq F_\alpha \frac{kf(1+T_r)}{e(f+1-k-r)} \quad (4.9)$$

for all $\mathbf{u}$ in the subspace spanned by the vectors in (4.5) is not less than $1 - \alpha$. The simultaneous confidence limits for all $\mathbf{u}'\boldsymbol{\tau}$ where $\mathbf{u}'$ belongs to $\mathbf{U}$ is thus provided by

$$u_1\hat{\tau}_1 + \dots + u_m\hat{\tau}_m \pm \sqrt{\left(F_\alpha \frac{kf(1+T_r)}{e(f+1-k-r)} \right)}. \quad (4.10)$$

The argument employed here is the same as that used by Tukey (1951), Scheffé (1953), Roy & Bose (1953), Hoel (1954) and others. It may be observed that the first attempt to obtain such simultaneous confidence intervals was due to Hotelling & Working (1929). The result (4.10) differs from that given by Hoel (1954) in that, in effect, he bases the confidence interval on the distribution of F on p , the number of variables $\mathbf{y}$, and $(f+1-p)$ degrees of freedom irrespective of the dimensionality of the subspace of vectors $\mathbf{u}'$ for which simultaneous confidence intervals are intended. Unless the factor $(1+T_r)$ in (4.10) is very large, which may happen when the specification (1.1) is not correct, in which case the procedure suggested is not valid, the confidence interval based on F , with k and $(f+1-k-r)$ degrees of freedom should be smaller in length when $k < p$.

5. AN ILLUSTRATIVE EXAMPLE

The results of the preceding sections show that the computations involved in tests of hypotheses and estimation of parameters are the same as in the method of least squares with correlated variables, although the distributions are slightly different due to the estimated dispersion matrix being used in the analysis instead of the true one. The computations can be

carried up to a certain stage mechanically as in the method of least squares with correlated variables. We may briefly review the methods to make this section self-contained for the use of those interested in the computational techniques.

Consider the quadratic form

$$(\mathbf{y} - \mathbf{A}\boldsymbol{\tau})' \mathbf{S}^{-1}(\mathbf{y} - \mathbf{A}\boldsymbol{\tau}) \quad (5.1)$$

with the estimated matrix $\mathbf{S}$ for the unknown dispersion matrix and obtain the normal equations

$$\mathbf{A}'\mathbf{S}^{-1}\mathbf{y} = \mathbf{A}'\mathbf{S}^{-1}\mathbf{A}\hat{\boldsymbol{\tau}}. \quad (5.2)$$

Find a solution $\hat{\boldsymbol{\tau}}$ and a matrix $\mathbf{C}$ of the variances and covariances of the estimates according to least squares theory assuming $\mathbf{S}$ to be true. The matrix $\mathbf{C}$ is the inverse of the matrix $\mathbf{A}'\mathbf{S}^{-1}\mathbf{A}$, and is unique if the rank of $\mathbf{A}$ is equal to the number of unknown parameters or a pseudo-inverse (which need not be unique) otherwise. The minimum value of the expression (5.1) is given by

$$\mathbf{y}'\mathbf{S}^{-1}\mathbf{y} - \mathbf{y}'\mathbf{S}^{-1}\mathbf{A}\hat{\boldsymbol{\tau}}. \quad (5.3)$$

The minimum value (5.3), $\hat{\boldsymbol{\tau}}$, a solution of the normal equations, and $\mathbf{C}$, which will be referred to as a least squares variance-covariance matrix, are used in obtaining confidence intervals and testing hypotheses as shown in the following illustrative example.

The example refers to observations on growth of ten rats under a control treatment considered in an earlier paper (Rao, 1958). The averages of 10 values representing increases in growth during 4 successive weeks are

$$y_0 = 24.5, \quad y_1 = 27.5, \quad y_2 = 24.1, \quad y_3 = 30.5$$

and the estimated dispersion matrix $\mathbf{S}$, within treatments, based on 24 D.F. (pooled from other treatment groups also) is

24.2625	1.7708	-2.3125	-3.1083
	25.3750	26.1042	14.3542
		43.6125	19.1250
			35.5417

In the notation of §1, $e = 10$ and $f = 24$. Let us examine whether a second degree polynomial in time can be fitted to the first differences of the growth curve or in other words whether

$$E(y_t) = \alpha + \beta t + \gamma t^2 \quad (t = 0, 1, 2, 3), \quad (5.4)$$

so that $r = 4 - 3 = 1$. The unknown parameters are α , β and γ and the matrix $\mathbf{A}$ of coefficients is

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix}$$

corresponding to the expectations for $t = 0, 1, 2, 3$.

The first step is to obtain the normal equations

$$\mathbf{A}'\mathbf{S}^{-1}\mathbf{y} = \mathbf{A}'\mathbf{S}^{-1}\mathbf{A}\hat{\boldsymbol{\tau}}$$

which involves the evaluation of the inverse of $\mathbf{S}$ and a triple product of matrices. A simple way of doing this is first to obtain a triangular reduction of the matrix $\mathbf{S}$, with the vector $\mathbf{y}$ and matrix $\mathbf{A}$ appended, by the square root or Gauss-Doolittle method (Dwyer, 1951; Rao, 1955).

Table 1. *Triangular reduction of matrix S, with y and A appended, for obtaining normal equations (square root method)*

Row no.	Matrix S				y	Matrix A		
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0.1	24.2625	1.7708	-2.3125	-3.1083	24.5	1	0	0
0.2	—	25.3750	26.1042	14.3542	27.5	1	1	1
0.3	—	—	43.6125	19.1250	24.1	1	2	4
0.4	—	—	—	35.5417	30.5	1	3	9
1.1	4.925698	0.359502	-0.469477	0.631037	4.973914	0.203017	0	0
1.2	—	5.024516	5.228957	2.901983	5.117282	0.184498	0.199024	0.199024
1.3	—	—	4.006257	0.912171	-0.080604	0.032591	0.239451	0.738670
1.4	—	—	—	5.088214	3.706994	0.110642	0.433160	1.522860

The rows (0.1) to (0.4) contain the matrix **S**, the vector **y** and the matrix **A**. We consider the *entire matrix* which is now in 8 columns and rows (0.1) to (0.4) and reduce it step by step by the square root method till the matrix **S** is reduced to the diagonal form. In this case only 4 steps are needed. The figures in columns (5) to (8) in the rows (1.1) to (1.4) containing the reduced matrices are used for further computations. We define the product of two columns as the sum of products of corresponding elements taken from rows (1.1) to (1.4). Thus

$$[5, 7] = (4.973914)(0) + (5.117282)(0.199024) + (-0.080604)(0.239451) + (3.706994)(0.433160) = 2.604883,$$

$$[5, 5] = (4.973914)^2 + \dots + (3.706994)^2 = 64.674697.$$

The normal equations are

$$\alpha[6, 6] + \beta[6, 7] + \gamma[6, 8] = [6, 5],$$

$$\alpha[7, 6] + \beta[7, 7] + \gamma[7, 8] = [7, 5],$$

$$\alpha[8, 6] + \beta[8, 7] + \gamma[8, 8] = [8, 5].$$

Computing these products and writing the equations in a form suitable for further computations we have

α	β	γ	Q
0.088560	0.092449	0.229286	2.361459
—	0.284575	0.876129	2.604883
—	—	2.904349	6.604173

where **Q** with elements $Q_\alpha, Q_\beta, Q_\gamma$, stands for the products of column (5) containing the vector **y** with the others. The reciprocal of the left-hand side matrix is

$$C = \begin{pmatrix} 22.867188 & -26.251429 & 6.113767 \\ & 79.444013 & -21.892726 \\ & & 6.465838 \end{pmatrix}$$

and by multiplying the rows of this matrix with the column $\mathbf{Q}$ we obtain estimates of α, β, γ as

$$\hat{\alpha} = 25.994127, \quad \hat{\beta} = 0.368153, \quad \hat{\gamma} = 0.110708.$$

The elements of the matrix $\mathbf{C}$ represent the variances and covariances of the estimates only when $\mathbf{S}$ is the true dispersion matrix. They have to be used slightly differently in expressing the precision of the estimates in the present case. The minimum value of $(\mathbf{y} - \mathbf{A}\boldsymbol{\tau})' \mathbf{S}^{-1}(\mathbf{y} - \mathbf{A}\boldsymbol{\tau})$ is $[5, 5] - \hat{\alpha}Q_{\alpha} - \hat{\beta}Q_{\beta} - \hat{\gamma}Q_{\gamma} = 1.6005$. If the object is to obtain this minimum value only then it is not necessary to obtain explicitly the solutions of the normal equations. The triangular reduction of the normal equations, say by the square root method, gives three values in the column for $\mathbf{Q}$, the sum of squares of which is equal to the expression $\hat{\alpha}Q_{\alpha} + \hat{\beta}Q_{\beta} + \hat{\gamma}Q_{\gamma}$. If estimates of parameters and their precisions are also required the method suggested above may be followed.

Test for specification. In the present problem, it is specified that the average differential growth curve of the population of rats is a second degree polynomial in time. To test whether this specification is adequate or not, we use the statistic (2.5, 2.9)

$$\frac{(f+1-r)e}{rf} \min_{\boldsymbol{\tau}} (\mathbf{y} - \mathbf{A}\boldsymbol{\tau})' \mathbf{S}^{-1}(\mathbf{y} - \mathbf{A}\boldsymbol{\tau}),$$

which is a variance ratio with degrees of freedom $r = 1$ and $(f+1-r) = 24$ in the present problem. The value of the statistic

$$\frac{24}{1} \frac{1.0}{24} (1.6005) = 16.005 \quad (5.5)$$

exceeds 4.25, the 5% point of F on 1 and 24 D.F., thus indicating inadequacy of the specification at that level of probability.

It may be observed that if the object is only to test the adequacy of specification (5.4), the alternative method of computing the matrix $\mathbf{B}$ and the quadratic form (2.4) suggested in §2 would be simpler in the present case. The matrix $\mathbf{A}'$ is

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{pmatrix}$$

and by inspection we observe that the matrix with rank $r = 1$, which is orthogonal to $\mathbf{A}'$ is

$$\mathbf{B} = \begin{pmatrix} -1 & 3 & -3 & 1 \end{pmatrix},$$

from which we compute

$$z_1 = (-1)y_0 + 3y_1 - 3y_2 + y_3 = 16.2, \\ \mathbf{BSB}' = 163.9081.$$

Since there is only one linear function z_1 , the appropriate statistic is Student's t . We can refer

$$t^2 = \frac{f+1-1}{1} T_1 = \frac{fe}{1f} \frac{z_1^2}{\mathbf{BSB}'} = 16.011 \quad (5.6)$$

to an F table with 1 and 24 D.F. This value agrees, to the degree of approximation expected, with that in (5.5) obtained by a different procedure. Although the second method is more direct, it does not provide estimates of the constants, which may be needed in actual practice. But if the sole object is a test for adequacy of the specification then the second method may be simpler.

Precision of estimates. In view of the fact that the assumed specification is not found to be adequate by the above test the problem of providing estimates of the parameters α, β, γ and

their precisions does not arise. We need to look for an alternative specification which will be supported by the data. But for purposes of illustration the computations laid out above have been used in the following, bearing in mind that the estimates and their precisions acquire a meaning only when the assigned specification is not rejected by a test.

The confidence interval for a parametric function $u_1 \tau_1 + \dots + u_m \tau_m$ is given by the formula (4.3). Obtain the least squares estimate

$$d = u_1 \hat{\tau}_1 + \dots + u_m \hat{\tau}_m$$

and least squares variance, $g = \mathbf{u}'\mathbf{C}\mathbf{u}$. The confidence interval is

$$d \pm t \left(\frac{(f-r)e}{gf(1+T_r)} \right)^{-\frac{1}{2}},$$

where

$$T_r = \frac{e}{f} \min (\mathbf{y} - \mathbf{A}\boldsymbol{\tau})' \mathbf{S}^{-1} (\mathbf{y} - \mathbf{A}\boldsymbol{\tau}).$$

In the present example, if a confidence interval for the parameter α is desired the formula is

$$\hat{\alpha} \pm t \left(\frac{(f-r)e}{c_{11}f(1+T_r)} \right)^{-\frac{1}{2}},$$

where c_{11} , the first diagonal element in the matrix $\mathbf{C}$, is the least squares variance of the estimate of α . Substituting the observed values we have the limits 25.9941 ± 4.1263 , corresponding to 95% probability.

Confidence bands to the differential growth curve. Assuming the specification to be true, let us consider the problem of obtaining confidence bands to the polynomial $\alpha + \beta t + \gamma t^2$ (for all values of the time). Although vectors of the form $(1, t, t^2)$ do not cover all the points of a three-dimensional space the probability statement (4.9) applicable to all combinations of α, β, γ is still true as an inequality. For an arbitrary value of time t the estimated polynomial value is

$$\alpha + \beta t + \gamma t^2 = 25.9941 + 0.3682t + 0.1107t^2$$

and the least squares variance of the estimate is the product

$$(1, t, t^2) \mathbf{C} \begin{pmatrix} 1 \\ t \\ t^2 \end{pmatrix} = 22.8672 - 52.5029t + 91.6715t^2 - 43.7855t^3 + 6.4658t^4.$$

The other values needed for an application of (4.10) are

$$k = 3, \quad f = 24, \quad e = 10, \quad 1 + T_r = 1.6669, \quad F_\alpha = 3.07, \quad \alpha = 0.05.$$

Substituting $t = 0, 1, 2, 3$ we obtain the following confidence limits:

	$t = 0$	$t = 1$	$t = 2$	$t = 3$
Lower limit	19.6600	19.8877	19.0383	20.6107
Mean value	25.9941	26.4729	27.1731	28.0947
Upper limit	32.3282	33.0581	35.3079	35.5787

From these values the limits to other values of t can be obtained by graduation (such as free hand smoothing) if necessary. The confidence bands are valid only when the specification adopted is supported by the data, for which a preliminary test is necessary.

REFERENCES

- DWYER, P. S. (1951). *Linear Computations*. New York: John Wiley and Sons.
- HOEL, P. G. (1954). Confidence bands for polynomial curves. *Ann. Math. Statist.* **25**, 534.
- HOTELLING, H. & WORKING, H. (1929). Applications of the theory of errors to the interpretation of trends. *J. Amer. Statist. Ass. (Suppl.)*, **24**, 73.
- HSU, P. L. (1938). Notes on Hotelling's generalized T . *Ann. Math. Statist.* **9**, 231.
- RAO, C. R. (1946*a*). On the linear combination of observations and the general theory of least squares. *Sankhyā*, **7**, 237.
- RAO, C. R. (1946*b*). Tests with discriminant functions in multivariate analysis. *Sankhyā*, **7**, 407.
- RAO, C. R. (1949). On some problems arising out of discrimination with multiple characters. *Sankhyā*, **9**, 343.
- RAO, C. R. (1952). *Advanced Statistical Methods in Biometric Research*. New York: John Wiley and Sons.
- RAO, C. R. (1955). Analysis of dispersion for multiply classified data with unequal numbers in cells. *Sankhyā*, **15**, 253.
- RAO, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics*, **14**, 1.
- ROY, S. N. & BOSE, R. C. (1953). Simultaneous confidence interval estimation. *Ann. Math. Statist.* **24**, 513.
- SCHEFFÉ, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika*, **40**, 87.
- TUKEY, J. W. (1951). Quick and dirty methods in statistics Part II. Simple analysis for standard designs. *Proc. Fifth Annual Convention Amer. Soc. for Quality Control*, p. 189.

TESTS OF SIGNIFICANCE IN CANONICAL ANALYSIS

By D. N. LAWLEY

Mathematical Institute, University of Edinburgh

SUMMARY. Results are obtained concerning the approximate distribution of canonical correlation coefficients. A slight improvement is suggested for the test of significance of residual canonical roots.

1. INTRODUCTION

In this paper our object is to investigate in some detail the approximate distribution of canonical correlation coefficients and the tests of significance associated with them. In particular we shall discuss the approximate test of significance proposed by Bartlett (1938, 1947) for the residual canonical roots when, let us say, the first k roots, corresponding to real relationships between two sets of variates, have been removed. Our approach involves some rather tedious and laborious algebra, but we have been unable to find any which is more elegant. No attempt has been made to justify the various expansions in series; they are presumed to be valid in an asymptotic sense.

2. ASSUMPTIONS AND NOTATION

Let two sets of variates containing respectively p and q variates be denoted by column vectors x_1 and x_2 , the corresponding row vectors being x'_1 and x'_2 . We shall be interested in the canonical relationships between x_1 and x_2 , and we shall begin by considering the true correlation case where all variates follow a multivariate normal distribution within each set is Without loss of generality, since at most only a linear transformation within each set is required, we may assume that their variances and covariances are given in matrix form by

$$\begin{aligned} E(x_1 x'_1) &= I_p, & E(x_2 x'_2) &= I_q, \\ E(x_1 x'_2) &= P, & E(x_2 x'_1) &= P', \end{aligned} \quad (1)$$

where I_p denotes the unit matrix of order p and where $P = [\rho_{ij}]$ is a matrix with p rows and q columns such that $\rho_{ij} = 0$ if $i \neq j$ and such that $\rho_{ii} = \rho_i$ is the i th canonical correlation coefficient in order of magnitude. We shall assume that the first k ($< p, < q$) of the ρ_i are distinct and positive and that the remainder are zero.

Corresponding to the above true values we shall suppose that the usual sample estimates of the variances and covariances, with n degrees of freedom, are given by

$$\begin{aligned} A_{11} &= I_p + X_{11}, & A_{22} &= I_q + X_{22}, \\ A_{12} &= P + X_{12}, & A_{21} &= P' + X_{21}. \end{aligned}$$

Thus the X -matrices represent deviations of sample estimates from expectation. The squares $r_1^2, \dots, r_q^2$ of the estimated canonical correlation coefficients are the latent roots of the matrix

$$A_{22}^{-1} A_{21} A_{11}^{-1} A_{12} = (I + X_{22})^{-1} (P' + X_{21}) (I + X_{11})^{-1} (P + X_{12}). \quad (2)$$

If $p < q$, the last $q - p$ of these roots are identically zero. The usual difficulty arises with regard to the correspondence of these roots with their population values. We shall assume, however, that they are arranged in order of magnitude and that with high probability r_i^2 then

corresponds with ρ_i^2 for $i = 1, 2, \dots, k$. It would be difficult to investigate the point with any degree of rigour, but it seems likely that, roughly speaking, we are justified in this assumption provided that the sample size is reasonably large and that the first k of the ρ_i are neither very near one another nor small.

Let us now expand the matrix product (2) in the form

$$(I - X_{22} + X_{22}^2 - \dots)(P' + X_{21})(I - X_{11} + X_{11}^2 - \dots)(P + X_{12}) = P'P + Y,$$

where $P'P$ is a diagonal matrix with elements $\rho_1^2, \dots, \rho_q^2$. The matrix Y may be expressed as

$$Y = Y^{(1)} + Y^{(2)} + Y^{(3)} + \dots,$$

$$\left. \begin{aligned} \text{where } Y^{(1)} &= X_{22}(I - P'P) - X_{33}, \\ Y^{(2)} &= -X_{22}^2(I - P'P) + X_{22}X_{33} + X_{31}X_{13}, \\ Y^{(3)} &= X_{22}^3(I - P'P) - X_{22}^2X_{33} - X_{22}X_{31}X_{13} - X_{31}X_{11}X_{13}, \text{ etc.,} \end{aligned} \right\} \quad (3)$$

$$\begin{aligned} \text{and where } X_{33} &= X_{22} - P'X_{12} - X_{21}P + P'X_{11}P, \\ X_{31} &= X'_{13} = X_{21} - P'X_{11}. \end{aligned}$$

In order to interpret the above notation let us introduce a vector x_3 of variates given by

$$x_3 = x_2 - P'x_1.$$

Thus x_3 represents the component of x_2 which is independent of x_1 . The variances and covariances involving x_3 are then given by

$$\left. \begin{aligned} E(x_3 x'_3) &= E(x_3 x'_2) = I - P'P, \\ E(x_3 x'_1) &= 0 \end{aligned} \right\} \quad (4)$$

The matrices X_{33} and X_{31} are to be interpreted as representing corresponding deviations of sample estimates from expectation.

3. VARIOUS EXPECTATIONS

We now require the expectations of elements and various products of elements of the matrices $Y^{(1)}, Y^{(2)}, Y^{(3)}$ and of Y . By the use of equations (3) we can express these in terms of the expectations of products of elements of the X -matrices. The latter can be evaluated by the usual combinatorial rules and by the use of the relations (1) and (4). The algebra is, however, extremely heavy, so we shall not reproduce it in full. Let us denote the typical elements of Y and of $Y^{(r)}$ by y_{ij} and $y_{ij}^{(r)}$ respectively ($i, j = 1, 2, \dots, q$). Then

$$\text{and we find, first, that } y_{ij} = y_{ij}^{(1)} + y_{ij}^{(2)} + y_{ij}^{(3)} + \dots,$$

$$E(y_{ii}) = (1 - \rho_i^2)(p - a_1 - \rho_i^2)/n + O(n^{-2}),$$

where

$$a_1 = \sum_{j=1}^k \rho_j^2.$$

For $i \neq j$ we have $E(y_{ij}) = 0$.

Proceeding further, for $i \neq j$, we have

$$\begin{aligned} E(y_{ij}^{(1)} y_{ji}^{(1)}) &= (1 - \rho_i^2)(1 - \rho_j^2)(\rho_i^2 + \rho_j^2)/n, \\ E(y_{ij}^{(1)} y_{ji}^{(2)}) &= -(1 - \rho_i^2)(1 - \rho_j^2)\{(q+2)\rho_i^2 + p\rho_j^2\}/n^2, \\ E(y_{ij}^{(1)} y_{ji}^{(3)}) &= (1 - \rho_i^2)(1 - \rho_j^2)\{(q-p+a_1+1)\rho_i^2 + (a_1-1)\rho_j^2 + \rho_i^4 + \rho_j^4 + 2\rho_i^2\rho_j^2\}/n^2, \\ E(y_{ij}^{(2)} y_{ji}^{(2)}) &= (1 - \rho_i^2)(1 - \rho_j^2)\{(p-2a_1+a_2) + (a_1-2)(\rho_i^2 + \rho_j^2) + 2(\rho_i^4 + \rho_j^4 + \rho_i^2\rho_j^2)\}/n^2, \end{aligned}$$

where

$$a_2 = \sum_{s=1}^k \rho_s^4.$$

We hence find that

$$E(y_{ij}y_{ji})/(1-\rho_i^2)(1-\rho_j^2) = (\rho_i^2 + \rho_j^2)/n + \{(p-2a_1+a_2) + (3a_1-2p+4)(\rho_i^2 + \rho_j^2) + 2(2\rho_i^4 + 2\rho_j^4 + 3\rho_i^2\rho_j^2)\}/n^2 + O(n^{-3}).$$

In a similar manner we obtain the following results:

$$\text{var}(y_{ii}) = 4\rho_i^2(1-\rho_i^2)^2/n + (1-\rho_i^2)(q-a_1)/n^2 + (1-\rho_i^2)^2\{(2p-q-2a_1+a_2) + 2(6a_1-4p-11)\rho_i^2 + 39\rho_i^4\}/n^2 + O(n^{-3}),$$

$$\kappa_3(y_{ii}) = 24\rho_i^2(1-\rho_i^2)^3(1-3\rho_i^2)/n^2 + O(n^{-3}),$$

$$\kappa_4(y_{ii}) = 96\rho_i^2(1-\rho_i^2)^4(1-2\rho_i^2)(2-13\rho_i^2)/n^3 + O(n^{-4}).$$

Also, for $i \neq j$, we have

$$\text{var}(y_{ij}) = (1-\rho_j^2)(\rho_i^2 + \rho_j^2 - 2\rho_i^2\rho_j^2)/n + O(n^{-2}),$$

$$\text{cov}(y_{ii}, y_{jj}) = (1-\rho_i^2)(1-\rho_j^2)\{-5(\rho_i^2 + \rho_j^2) + 4(\rho_i^4 + \rho_j^4) + 5\rho_i^2\rho_j^2\}/n^2 + O(n^{-3}),$$

$$\text{cov}\{y_{ii}, (y_{ij}y_{ji})\} = (1-\rho_i^2)^2(1-\rho_j^2)(1+5\rho_i^2+\rho_j^2-8\rho_i^4-\rho_j^4-9\rho_i^2\rho_j^2)/n^2 - (1-\rho_i^2)(1-\rho_j^2)^2/n^2 + O(n^{-3}).$$

4. APPROXIMATE DISTRIBUTION OF THE FIRST k ROOTS

We now make use of an expression for the first k latent roots of the matrix $P'P + Y$. A similar expansion was derived in a previous paper (Lawley, 1956*a*) in which we were concerned with the internal analysis of a single set of variates. The s th root of the matrix ($s = 1, 2, \dots, k$) is given by

$$r_s^2 = \rho_s^2 + y_{ss} + \sum_i' \frac{y_{si}y_{is}}{\rho_s^2 - \rho_i^2} - y_{ss} \sum_i' \frac{y_{si}y_{is}}{(\rho_s^2 - \rho_i^2)^2} + \sum_{i,j}' \frac{y_{si}y_{ij}y_{js}}{(\rho_s^2 - \rho_i^2)(\rho_s^2 - \rho_j^2)} + O(y^4),$$

where Σ' indicates summation over all values from 1 to q with the exception of s . The validity of the expansion depends upon the assumption that the y -variates, which are of order $1/\sqrt{n}$, are small compared with the quantities $\rho_s^2 - \rho_i^2$.

By means of this expansion and of the results of the preceding section we are able to evaluate the first four cumulants of r_s^2 and hence those of r_s . We find that

$$2\rho_s E(r_s - \rho_s) = (1-\rho_s^2) \left\{ p+q-2-\rho_s^2 + 2(1-\rho_s^2) \sum_i' \left(\frac{\rho_i^2}{\rho_s^2 - \rho_i^2} \right) \right\} / n + O(n^{-2}),$$

$$\text{var}(r_s) = (1-\rho_s^2)^2/n + O(n^{-2}).$$

$$\kappa_3(r_s) = -6\rho_s(1-\rho_s^2)^3/n^2 + O(n^{-3}),$$

$$\kappa_4(r_s) = -6(1-\rho_s^2)^4(1-12\rho_s^2)/n^3 + O(n^{-4}).$$

In general, it seems useless to seek a better approximation to the variance of r_s , since the term of order n^{-2} is extremely cumbersome. The variance depends to a considerable extent not only on the value of ρ_s but also on the differences $\rho_s^2 - \rho_i^2$. It is, however, worth considering the case where $k = 1$, i.e. where ρ_1 is the only non-zero canonical correlation coefficient. More generally, for $k > 1$, we can consider r_k under the somewhat restrictive assumption that the first $k-1$ of the ρ_i are near unity while ρ_k is moderate. For these cases, writing ρ in place of ρ_k we have

$$E(r_k) \approx \rho + \frac{1}{2}(1-\rho^2)(m\rho^{-1}-\rho)/n,$$

$$\text{var}(r_k) \approx (1-\rho^2)^2/n + \frac{1}{2}(1-\rho^2)^2\{11\rho^2 - 2(m-k+1) - m\rho^{-2}\}/n^2,$$

where

$$m = p+q-2k.$$

It is natural to apply the usual z -transformation in order to normalize the distributions of the canonical correlation coefficients; but in general this is unfortunately of little help

since the expectations of all other terms are of order less than n^{-2} and may be neglected. We find that

$$\begin{aligned} E(\text{tr } Y_{22}^2) &= (q-k) \{(q-k)(p-2a_1+a_2) + (p-a_1)(p-a_1+1)\}/n^2, \\ E(\text{tr } Y_{22} Y_{21} D^{-1} Y_{12}) &= (q-k) \{(q-k)(k-2a_1+a_2) + (k-a_1)(p-a_1+1)\}/n^2, \\ E\{\text{tr } (Y_{21} D^{-1} Y_{12})^2\} &= (q-k) \{(q-k)(k-2a_1+a_2) + (k-a_1)(k-a_1+1)\}/n^2, \end{aligned}$$

and hence that

$$E(\text{tr } Z^2) = (p-k)(q-k)(p+q-2k+1)/n^2.$$

The expectations of $\text{tr } Z^3$, $\text{tr } Z^4$, etc., are all of order less than n^{-2} and may be neglected. Hence, finally the expectation of (7) is found to be

$$(p-k)(q-k) \{1/n + \frac{1}{2}(p+q+2k-2b+1)/n^2\} + O(n^{-3}).$$

Thus, supposing that the values of $\rho_1, \dots, \rho_k$ were known, the correct multiplying factor for the criterion (6) would be

$$n-k - \frac{1}{2}(p+q+1) + \sum_{j=1}^k (1/\rho_j^2).$$

This would not only make the expectation of the criterion equal to $(p-k)(q-k) + O(n^{-2})$, but would make the other moments equal to those of χ^2 apart from errors of order n^{-2} , by a result established elsewhere (Lawley, 1956b).

If $\rho_1, \dots, \rho_k$ are all fairly near unity, the multiplying factor is approximately $n - \frac{1}{2}(p+q+1)$, which is the value suggested by Bartlett. Otherwise an improvement should, presumably, be obtained by taking the factor as

$$n-k - \frac{1}{2}(p+q+1) + \sum_{j=1}^k (1/\rho_j^2), \quad (8)$$

though we cannot claim this to be very firmly established.

6. THE CASE OF FIXED INDEPENDENT VARIATES

Let us now suppose that x_1 represents a set of fixed independent variates and x_2 the dependent variates. Then the canonical analysis is to be regarded as a regression analysis of x_2 on x_1 . For given x_1 the variates x_2 are assumed to be normally distributed with constant variances and covariances and with mean values given by the vector $B'x_1$, where $B = [\beta_{ij}]$ is a $p \times q$ matrix whose elements represent coefficients of linear regression. Without loss of generality we may assume that the variates are chosen in such a way that $\beta_{ij} = 0$ if $i \neq j$ and that $\beta_{ii} = \beta_i$, where $\beta_1, \dots, \beta_k$ are positive and arranged in order of magnitude while the remaining β_i are zero. We may also assume that, for given x_1 , the variates x_2 have unit variances and zero covariances. These assumptions require a change of notation.

In considering sample values it is now convenient to use sums of squares and products (with n degrees of freedom). These form matrices which we denote by $C_{11}, C_{12} = C'_{21}$ and C_{22} . It will be convenient to consider the matrix

$$(C_{22} - C_{21} C_{11}^{-1} C_{12})^{-1} C_{21} C_{11}^{-1} C_{12}, \quad (9)$$

which has for its latent roots the quantities $r_i^2/(1-r_i^2)$, where $i = 1, 2, \dots, q$.

Estimates, with $m = n - p$ degrees of freedom, of the variances and covariances of the x_2 (for given x_1) are provided by the elements of the matrix

$$\frac{1}{m} (C_{22} - C_{21} C_{11}^{-1} C_{12}) = I_q + V,$$

where the elements of V represent deviations of these estimates from expectation. To simplify matters we may choose the x_1 , again without loss of generality, in such a way that $C_{11} = mI_p$. Then we may write

$$\frac{1}{m}C_{12} = C_{11}^{-1}C_{12} = B + X,$$

where the elements of the $p \times q$ matrix X represent deviations of estimated regression coefficients from expectation. They constitute a set of pq independent normal variates each having zero mean and variance $1/m$ and they are distributed independently of the elements of V .

The matrix given by (9) may now be expressed in the form

$$(I + V)^{-1}(B' + X')(B + X) = (I - V + V^2 - \dots)(B' + X')(B + X) = B'B + Y,$$

where $B'B$ is a diagonal matrix with elements $\beta_i^2 (i = 1, 2, \dots, q)$. The first k roots of (9) provide estimates of $\beta_1^2, \dots, \beta_k^2$, while the remainder enable us to test the hypothesis that $\beta_{k+1}, \dots, \beta_q$ are all zero. The matrix Y may be expanded as before in the form

$$Y = Y^{(1)} + Y^{(2)} + Y^{(3)} + \dots,$$

but now

$$\begin{aligned} Y^{(1)} &= B'X + X'B - VB'B, \\ Y^{(2)} &= X'X - VB'X - VX'B + V^2B'B, \\ Y^{(3)} &= -VY^{(2)}, \\ Y^{(4)} &= V^2Y^{(2)}, \text{ etc.} \end{aligned}$$

Let us partition Y in the same manner as the Y of the preceding section. Then the $q - k$ smallest roots of $B'B + Y$ are approximately the same as those of the matrix Z , where Z is as given by (5), provided that we now define D as the diagonal matrix with elements $\beta_1^2, \dots, \beta_k^2$. The latent roots of $I + Z$ are approximately equal to the quantities $1/(1 - r_i^2)$, for $i = k + 1, \dots, q$. Hence the criterion (6) may in this case be expressed as

$$\log_e |I + Z| = \text{tr}(Z) - \frac{1}{2} \text{tr}(Z^2) + \frac{1}{3} \text{tr}(Z^3) - \dots \quad (10)$$

In order to evaluate the expectation of this expression it is necessary to recalculate the expectations of the various products of elements of Y . This is however a comparatively simple process. Assuming throughout that $\beta_s = 0 (k < s \leq q)$ and neglecting quantities of order m^{-3} we find that

$$\begin{aligned} E(y_{ss}) &= p/m + p(q+1)/m^2, \\ E(y_{ss}^2) &= (p^2 + 2p + c)/m^2, \end{aligned}$$

where

$$c = \sum_{j=1}^k \beta_j^2.$$

For $i \neq s$ we have

$$\begin{aligned} E(y_{si}y_{is}) &= \beta_i^2/m + \{p + (p + 2q + 2)\beta_i^2\}/m^2, \\ E(y_{ss}y_{si}y_{is}) &= \beta_i^2(p + 2 + \beta_i^2)/m^2, \\ E(y_{si}y_{ii}y_{is}) &= \beta_i^2\{(p + 4) + (q + 4)\beta_i^2\}/m^2, \\ E(y_{si}^2y_{is}^2) &= \beta_i^4(3 + \beta_i^2)/m^2, \\ E(y_{si}y_{ii}^2y_{is}) &= 2\beta_i^4(2 + \beta_i^2)/m^2; \end{aligned}$$

and, for i, j, s all unequal,

$$\begin{aligned} E(y_{si}y_{ij}y_{js}) &= (\beta_i^2 + \beta_j^2 + 2\beta_i^2\beta_j^2)/m^2, \\ E(y_{si}y_{ij}y_{ji}y_{is}) &= \beta_i^2(\beta_i^2 + \beta_j^2 + \beta_i^2\beta_j^2)/m^2, \\ E(y_{si}y_{is}y_{sj}y_{js}) &= \beta_i^2\beta_j^2/m^2. \end{aligned}$$

We then have

$$\begin{aligned} E(\text{tr } Y_{22}) &= p(q-k)/m + p(q-k)(q+1)/m^2, \\ E(\text{tr } Y_{21} D^{-1} Y_{12}) &= k(q-k)/m + (q-k)\{pd + k(p+2q+2)\}/m^2, \\ E(\text{tr } Y_{22} Y_{21} D^{-2} Y_{12}) &= (q-k)\{(p+q-k+1)d + k\}/m^2, \\ E(\text{tr } Y_{21} D^{-1} Y_{11} D^{-1} Y_{12}) &= (q-k)\{(p+2k+2)d + k(q+2k+2)\}/m^2, \\ E(\text{tr } Y_{21} D^{-1} Y_{11} D^{-1} Y_{11} D^{-1} Y_{12}) &= (q-k)(k+1)(k+2d)/m^2, \\ E(\text{tr } Y_{21} D^{-1} Y_{12} Y_{21} D^{-2} Y_{12}) &= (q-k)\{(q+1)d + k\}/m^2, \end{aligned}$$

where

$$d = \sum_{j=1}^k (1/\beta_j^2).$$

Hence

$$E(\text{tr } Z) = (p-k)(q-k)\{1/m + (q-k+1-d)/m^2\}.$$

We have also

$$\begin{aligned} E(\text{tr } Y_{22}^2) &= (q-k)\{p(p+q-k+1) + c\}/m^2, \\ E(\text{tr } Y_{22} Y_{21} D^{-1} Y_{12}) &= (q-k)\{k(p+q-k+1) + c\}/m^2, \\ E\{\text{tr } (Y_{21} D^{-1} Y_{12})^2\} &= (q-k)\{k(q+1) + c\}/m^2; \end{aligned}$$

and thus

$$E(\text{tr } Z^2) = (p-k)(q-k)(p+q-2k+1)/m^2.$$

Hence, finally, the expectation of (10) is found to be

$$\begin{aligned} (p-k)(q-k)\{1/m + \frac{1}{2}(q-p-2d+1)/m^2\} + O(m^{-3}) \\ = (p-k)(q-k)\{1/n + \frac{1}{2}(p+q-2d+1)/n^2\} + O(n^{-3}), \end{aligned}$$

putting $m = n - p$.

The appropriate multiplying factor for the criterion (6) is

$$n - \frac{1}{2}(p+q+1) + \sum_{j=1}^k (1/\beta_j^2).$$

If $\beta_1, \dots, \beta_k$ are all large the last term may be neglected. If, on the other hand, we replace each β_j^2 by its estimate $r_j^2/(1-r_j^2)$, the multiplying factor will again be given by expression (8). We are thus led to much the same conclusion as in the true correlation case.

REFERENCES

- BARTLETT, M. S. (1938). Further aspects of the theory of multiple regression. *Proc. Camb. Phil. Soc.* **34**, 33-40.
 BARTLETT, M. S. (1947). Multivariate analysis. *J.R. Statist. Soc. (Suppl.)*, **9**, 176-90.
 LAWLEY, D. N. (1956a). Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika*, **43**, 128-36.
 LAWLEY, D. N. (1956b). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, **43**, 295-303.

THE ESTIMATION OF PARAMETERS IN SYSTEMS OF STOCHASTIC DIFFERENTIAL EQUATIONS

BY A. W. PHILLIPS

London School of Economics

1. SUMMARY

In many fields of study there occur systems of lagged dependences which can be described approximately by systems of linear stationary stochastic differential equations. In this paper we consider the estimation of parameters in such systems, using an approach similar to that of Bartlett (1946, 1955) and Quenouille (1957), and discuss conditions for identification of the separate behaviour relationships of a system. Little is known about the small-sample properties of estimates in systems of this kind, but sampling experiments are being carried out and the results will be published later.

2. SYSTEMS OF LINEAR STOCHASTIC LAGGED DEPENDENCES

2. SYSTEMS OF LINEAR STOCHASTIC LAGGED DEPENDENCES

A system of linear stochastic lagged dependences between n variables $x_i(t)$ which are functions of continuous time can be described by the system of convolution integrals

$$x_i(t) = \sum_{j \neq i} \int_0^\infty w_{ij}(h) x_j(t-h) dh + \xi_i(t) \quad (i, j = 1, 2, \dots, n), \quad (1)$$

in which each of the weighting functions $w_{ij}(h)$ specifies the magnitude of the dependence of the value taken by x_i at time t on the values taken by x_j at times $t-h$, and the $\xi_i(t)$ are stochastic disturbance terms. The Laplace transform of the weighting function of a lagged dependence is called the transfer function of the dependence. If all the transfer functions of the system (1) are rational functions, i.e. if

$$\int_0^\infty e^{-sh} w_{ij}(h) dh = \frac{\beta_{ij}(s)}{\alpha_{ij}(s)}, \quad (2)$$

where $\beta_{ij}(s)$ and $\alpha_{ij}(s)$ are polynomials of finite degree in s , (1) can be written in the form of differential equations

$$x_i(t) = \sum_{j \neq i} \frac{\beta_{ij}(D)}{\alpha_{ij}(D)} x_j(t) + \xi_i(t) \quad (i, j = 1, 2, \dots, n) \quad (3)$$

in which D is the differential operator d/dt . For the i th equation in (1) is then a particular integral of the i th equation in (3), and if the system is stable and the initial conditions are at $t = 0$, the i th equation in (3).

Let $Q_{ij}(D)$ be the elements of $Q(D)$ whose elements $Q_{ij}(D)$ are given by

$$Q_{ij}(D) = \begin{cases} -\frac{\beta_{ij}(D)}{\alpha_{ij}(D)} & \text{if } j \neq i, \\ 1 & \text{if } j = i \end{cases} \quad (4)$$

and let $x_i(t)$ and $\xi_i(t)$ be vectors with elements $x_i(t)$ and $\xi_i(t)$. Then (3) may be written

$$Q(D)x(t) = \xi(t).$$

In many fields of study a system like (3) may be used to represent the behaviour of a number of separate units which form an interdependent system. Thus the equations may represent the behaviour of units of physical equipment, as in control engineering, or of groups of persons or firms, as in economics, or of animals or organisms, as in ecological or biological studies. In describing real behaviour systems of this sort it is often reasonable to assume that the behaviour units cannot respond to change instantaneously. This is equivalent to the assumption that the weighting functions $w_{ij}(h)$ in (1) do not contain delta-function impulse terms,* which in turn implies that the numerator polynomial $\beta_{ij}(s)$ of each transfer function is of lower degree than the denominator polynomial $\alpha_{ij}(s)$. In this paper we shall consider only systems in which the numerator of each transfer function is of lower degree than the denominator. It will be seen later that this restriction plays an important part in the identification of the separate behaviour equations of a system.

3. SPECIFICATION OF THE STOCHASTIC DISTURBANCES $\xi(t)$

We shall consider only systems which are stable and in which the disturbances $\xi(t)$ are stationary stochastic variables whose covariance functions possess rational two-sided Laplace transforms. It will be convenient to assume that the disturbances $\xi(t)$ are generated from another set of stationary stochastic variables $\zeta_k(t)$ by the transformation

$$\xi_i(t) = \sum_{k=1}^n \int_0^\infty r_{ik}(h) \zeta_k(t-h) dh \quad (i = 1, 2, \dots, n) \quad (6)$$

in which the functions $r_{ik}(h)$ have Laplace transforms $R_{ik}(s)$ of the form

$$R_{ik}(s) = \frac{\eta_{ik}(s)}{\gamma_{ik}(s)}, \quad (7)$$

where $\eta_{ik}(s)$ and $\gamma_{ik}(s)$ are polynomials in s , the degree of $\eta_{ik}(s)$ being lower than that of $\gamma_{ik}(s)$. Equations (6) can then be written as the set of differential equations

$$\xi(t) = R(D) \zeta(t), \quad (8)$$

where $R(D)$ is a matrix with elements $R_{ik}(D)$ and $\zeta(t)$ is a column vector with elements $\zeta_k(t)$. In order that the variables $\xi(t)$ be stationary it is necessary that all the roots of the equations $\gamma_{ik}(s) = 0$ have negative real parts.

We define the following covariance functions;

$$\left. \begin{aligned} \vartheta(\tau) &= E\{\xi(t) \xi'(t-\tau)\}, \\ \psi(\tau) &= E\{\xi(t) \zeta'(t-\tau)\}, \\ \lambda(\tau) &= E\{\zeta(t) \xi'(t-\tau)\} \end{aligned} \right\} \quad (9)$$

* The δ -function or unit impulse function, $\delta(h)$ is defined by the equations

$$\delta(h) = \begin{cases} 0 & \text{if } h \neq 0 \\ \infty & \text{if } h = 0 \end{cases} \quad \text{and} \quad \int_{-\infty}^{\infty} \delta(h) dh = 1.$$

(See James, Nichols & Phillips (1947), chapter 2, especially § 2.4.) The Laplace transform of $\delta(h)$ is clearly unity. If $\beta_{ij}(s)$ were not of lower degree than $\alpha_{ij}(s)$ the partial fraction expansion of $\beta_{ij}(s)/\alpha_{ij}(s)$ would contain a constant term and so its inverse transform $w_{ij}(h)$ would contain a δ -function impulse term.

and denote their Laplace transforms by $\theta(s)$, $\Psi(s)$ and $\Lambda(s)$. Thus

$$\left. \begin{aligned} \theta(s) &= \int_{-\infty}^{\infty} e^{-s\tau} \vartheta(\tau) d\tau, \\ \Psi(s) &= \int_{-\infty}^{\infty} e^{-s\tau} \psi(\tau) d\tau, \\ \Lambda(s) &= \int_{-\infty}^{\infty} e^{-s\tau} \lambda(\tau) d\tau. \end{aligned} \right\} \quad (10)$$

It will now be shown that $\theta(s)$ is given by

$$\theta(s) = R(s)\Psi(s)R'(-s). \quad (11)$$

The functions $r_{ik}(h)$ in (6) may be regarded as defined but zero for $h < 0$, and the lower limit of integration extended to $-\infty$. In matrix notation (6) then becomes

$$\xi(t) = \int_{-\infty}^{\infty} r(h)\xi(t-h)dh, \quad (12)$$

$r(h)$ being the matrix with elements $r_{ik}(h)$. Post-multiplying (12) by $\xi'(t-\tau)$ and taking expectations we obtain using (9),

$$\vartheta(\tau) = \int_{-\infty}^{\infty} r(h)\lambda(\tau-h)dh. \quad (13)$$

Taking the Laplace transform of (13) and remembering that the transform of the convolution of two functions is the product of their transforms* we have

$$\theta(s) = R(s)\Lambda(s). \quad (14)$$

Also, postmultiplying (12) by $\xi'(t+\tau)$ and taking expectations we find, using (9)

$$\lambda'(\tau) = \int_{-\infty}^{\infty} r(h)\psi'(\tau+h)dh. \quad (15)$$

On making the substitution $h_1 = -h$ this becomes

$$\lambda'(\tau) = \int_{-\infty}^{\infty} r(-h_1)\psi'(\tau-h_1)dh_1 \quad (16)$$

and it follows that

$$\Lambda'(s) = R(-s)\Psi'(s). \quad (17)$$

Substituting the transpose of (17) in (14) we obtain (11).

We now postulate that $\Psi(s)$ be identically the unit matrix,

$$\Psi(s) = I. \quad (18)$$

Substitution of (18) in (11) then gives

$$\theta(s) = R(s)R'(-s). \quad (19)$$

Since the spectral density and cross-spectral density functions of a set of variables can be obtained from the two-sided Laplace transforms of their covariance functions by replacing

* See Widder (1946), p. 258, theorem 16b. The conditions for the theorem are satisfied as a result of the stationarity of $\zeta(t)$ and the conditions that the roots of the equations $\gamma_{ik}(s) = 0$ have negative real parts.

the complex number s by the pure imaginary $i\omega$ (the Laplace transform then reducing to a Fourier transform), the spectral densities of the variables $\zeta(t)$ are constants for all frequencies ω and their cross-spectral densities are identically zero. The matrix $\psi(\tau)$ of their covariance functions is diagonal, the leading elements being δ -function impulses. Since this implies that the variances are infinite, the variables $\zeta(t)$ are not physically realizable.* However, (6) remains a valid mathematical model for the generation of the physically realizable variables $\xi(t)$, whose spectral density and cross-spectral density functions, given by $\theta(i\omega)$, are proper rational functions of $i\omega$ as a result of the restriction that $\eta_{ik}(s)$ in (7) be of lower degree than $\gamma_{ik}(s)$.

4. COVARIANCE FUNCTIONS OF THE SYSTEM VARIABLES $x(t)$

Substituting (8) in (5) we have

$$Q(D)x(t) = R(D)\zeta(t). \quad (20)$$

Let $P(D)$ be a diagonal matrix whose i th leading element is the lowest common multiple of the denominators of the elements in the i th row of $Q(D)$ and in the i th row of $R(D)$, i.e. the lowest common multiple of the polynomials $\alpha_{ij}(D)$, $j \neq i$, in (4) and $\gamma_{ik}(D)$, $k = 1, 2, \dots, n$ in (7), and let $F(D)$ and $G(D)$ be matrices defined by

$$F(D) = P(D)Q(D) \quad (21)$$

$$\text{and} \quad G(D) = P(D)R(D). \quad (22)$$

Then premultiplying (20) by $P(D)$ we obtain

$$F(D)x(t) = G(D)\zeta(t). \quad (23)$$

The elements of $F(D)$ and $G(D)$ are polynomials in D , the diagonal element in any row of $F(D)$ being of higher degree than the non-diagonal elements in the same row and also of higher degree than the elements in the same row of $G(D)$. Thus the determinant $|F(D)|$ cannot vanish identically, and we may write (23) as

$$x(t) = F^{-1}(D)G(D)\zeta(t). \quad (24)$$

Let $\phi(\tau)$ be the matrix of covariance functions of the system variables $x(t)$, so that

$$\phi(\tau) = E\{x(t)x'(t-\tau)\} \quad (25)$$

and let $\Phi(s)$ be the Laplace transform of $\phi(\tau)$. Then, by a similar argument to that used in §3, we have

$$\Phi(s) = F^{-1}(s)G(s)\Psi(s)G'(-s)F'^{-1}(-s) \quad (26)$$

$$\text{or, since } \Psi(s) = I, \quad \Phi(s) = F^{-1}(s)G(s)G'(-s)F'^{-1}(-s). \quad (27)$$

If we define the polynomial $p(s)$ by

$$\begin{aligned} p(s) &= |F(s)| \\ &= \prod_{r=1}^m (s - \lambda_r) \end{aligned} \quad (28)$$

* In engineering literature the variables $\zeta(t)$ are called independent white noise sources. See Laning & Battin (1956), pp. 136-44 and Miller (1956), pp. 282-4.

and write $F^a(s)$ for the adjoint of $F(s)$, (27) becomes

$$\Phi(s) = \frac{F^a(s) G(s) G'(-s) F'^a(-s)}{p(s) p(-s)}. \quad (29)$$

For simplicity we shall consider in this paper only cases in which the roots λ_r , $r = 1, 2, \dots, m$, of the equation $p(s) = 0$ are all distinct, and shall also assume that $p(s)$ does not contain any factor which is common to all the elements of $F^a(s)$. Expanding the right-hand side of (29) in partial fractions we then obtain

$$\Phi(s) = \sum_{r=1}^m \frac{K_r}{s - \lambda_r} + \sum_{r=1}^m \frac{K'_r}{-s - \lambda_r}, \quad (30)$$

where

$$K_r = \frac{(s - \lambda_r) F^a(s) G(s) G'(-s) F'^a(-s)}{p(s) p(-s)} \Big|_{s=\lambda_r}. \quad (31)$$

For a simple root λ_r the matrix $F(\lambda_r)$ is necessarily simply degenerate. The adjoint matrix $F^a(\lambda_r)$ is of unit rank and can be written

$$F^a(\lambda_r) = k_r \kappa_r, \quad (32)$$

where k_r is a column vector which satisfies the equation

$$F(\lambda_r) k_r = 0 \quad (33)$$

and κ_r is a row vector which satisfies

$$\kappa_r F(\lambda_r) = 0. \quad (34)$$

(See Frazer, Duncan & Collar (1950) 5.6.) Let l_r be a row vector defined by

$$l_r = \frac{(s - \lambda_r) \kappa_r G(s) G'(-s) F'^a(-s)}{p(s) p(-s)} \Big|_{s=\lambda_r}. \quad (35)$$

Then substituting (32) in (31) and using (35) we have

$$K_r = k_r l_r. \quad (36)$$

Thus the matrix K_r is of unit rank, with columns proportional to the vector k_r . It follows from (36) and (33) that

$$F(\lambda_r) K_r = 0 \quad (r = 1, 2, \dots, m), \quad (37)$$

a result which could also have been obtained by premultiplying (31) by $F(s)$ and putting $s = \lambda_r$.

The matrix $\phi(\tau)$ of the system covariance functions is obtained by taking the inverse transform of (30). Formally, (30) does not have a unique inverse. However, for a stationary system the roots λ_r are negative or have negative real parts and $\phi(\tau) \rightarrow 0$ as $\tau \rightarrow \pm \infty$. It can readily be verified that the only inverse which is consistent with these conditions is that obtained by taking the first sum in (30) to be the one-sided transform of $\phi(\tau)$ for $\tau \geq 0$ and the second sum to be the one-sided transform for $\tau < 0$. The inverse of the first sum in (30) then gives

$$\phi(\tau) = \sum_{r=1}^m K_r e^{\lambda_r \tau} \quad (\tau \geq 0) \quad (38)$$

while from the inverse of the second sum we see that $\phi(-\tau) = \phi'(\tau)$, a fact which is also obvious from (25).

5. CALCULATION OF $\Phi(s)$ FROM GIVEN POINTS ON $\phi(\tau)$

Let ϕ_τ be the matrix of discrete functions consisting of points at unit intervals of τ on the continuous functions $\phi(\tau)$, i.e. let

$$\phi_\tau = \phi(\tau) \quad \text{for } (\tau = \dots, -2, -1, 0, 1, 2, \dots), \quad (39)$$

and let $M(z)$ be the one-sided generating function defined by

$$M(z) = \sum_{\tau=0}^{\infty} \phi_\tau z^\tau. \quad (40)$$

Then from (38), (39) and (40) we find that

$$M(z) = \sum_{r=1}^m \frac{K_r}{1 - \mu_r z}, \quad (41)$$

where

$$\mu_r = e^{\lambda_r}. \quad (42)$$

We define the polynomial

$$\Delta(z) = \prod_{r=1}^m (1 - \mu_r z), \quad (43)$$

and seek matrices $U(z)$ and $V(z)$, whose elements are polynomials in z , such that

$$M(z) = U^{-1}(z) V(z). \quad (44)$$

If (44) is to hold we require that

$$|U(z)| = \Delta(z) \quad (45)$$

and that

$$U(z) \sum_{r=1}^m \frac{K_r}{1 - \mu_r z} = V(z). \quad (46)$$

Multiplying (46) by $1 - \mu_q z$ we obtain

$$U(z) \left\{ K_q + (1 - \mu_q z) \sum_{r \neq q} \frac{K_r}{1 - \mu_r z} \right\} = (1 - \mu_q z) V(z) \quad (47)$$

and on putting $z = 1/\mu_q$, $q = 1, 2, \dots, m$, this gives

$$U \left(\frac{1}{\mu_q} \right) K_q = 0 \quad (q = 1, 2, \dots, m). \quad (48)$$

Since K_q is of unit rank, (48) provides m scalar equations in the coefficients of the polynomials occurring in each row of $U(z)$. If $U(z)$ is given the form

$$U(z) = I + U_1 z + U_2 z^2 + \dots + U_c z^c, \quad (49)$$

where $U_1, U_2, \dots, U_c$ are $n \times n$ matrices whose elements are constants, each row of $U(z)$ will contain cn coefficients of powers of z (excluding the constant terms, which are unity or zero). It can be shown that if c is given a suitable value and $cn - m$ coefficients in each row of $U(z)$ are put equal to zero, equations (48) suffice to determine the remaining m coefficients in each row of $U(z)$, provided the positions of the zeros are chosen in such a way that $|U(z)|$ is of degree m . It will always be possible to find positions for the zeros such that this condition is satisfied, and equation (45) will then also be satisfied. If m is an integral multiple of n it will usually, but not always, be possible to take c equal to m/n , in which case $cn - m = 0$ and none of the coefficients is put equal to zero.

When $U(z)$ has been determined $V(z)$ can be found from (46). It can be seen from (48) that $1 - \mu_r z$ is a factor of $U(z) K_r$ and it follows from this and (46) that the degree of $V(z)$ is at least one less than that of $U(z)$. Thus $V(z)$ is of the form

$$V(z) = V_0 + V_1 z + \dots + V_{c-1} z^{c-1}, \quad (50)$$

where $V_0, V_1, \dots, V_{c-1}$ are matrices whose elements are constants.

Premultiplying (44) by $U(z)$ and writing the matrices in expanded form, we have

$$(I + U_1 z + \dots + U_c z^c)(\phi_0 + \phi_1 z + \dots) = V_0 + V_1 z + \dots + V_{c-1} z^{c-1}. \quad (51)$$

Equating the coefficients of $z^c, z^{c+1}, \dots$, in (51) we obtain the difference equation

$$\phi_\tau + U_1 \phi_{\tau-1} + \dots + U_c \phi_{\tau-c} = 0 \quad (\tau \geq c), \quad (52)$$

from which the matrices $U_1, U_2, \dots, U_c$ can be calculated if ϕ_τ is given for a sufficient number of values of τ . Equating the constant terms in (51) we have

$$V_0 = \phi_0 \quad (53)$$

and equating the coefficients of $z, z^2, \dots, z^{c-1}$ we find

$$V_l = \phi_l + \sum_{k=1}^l U_k \phi_{l-k} \quad (l = 1, 2, \dots, c-1), \quad (54)$$

from which the matrices $V_0, V_1, \dots, V_{c-1}$ can be calculated.

Thus given a sufficient number of points at equal intervals on the system covariance functions $\phi(\tau)$ it is possible to calculate $U(z)$ and $V(z)$. $M(z)$ is then given by (44), and can be expanded in partial fractions to give (41), from which the matrices K_r and the scalars μ_r are obtained immediately. The roots λ_r of the equation $p(s) = 0$ can then be calculated from

$$\lambda_r = \log \mu_r, \quad (55)$$

which follows from (42), and the matrix $\Phi(s)$ can then be found from the K_r and λ_r using (30).

6. CALCULATION OF $Q(s)$ AND $\theta(s)$ FROM $\Phi(s)$ WHEN ALL EQUATIONS ARE OF THE SAME ORDER

Consider an n variable system defined by (20) or (23), in which all the scalar equations of $F(D)x(t) = G(D)\zeta(t)$ are of the same order u , and which cannot be reduced to an equivalent system in which any equation is of lower order. The elements on the leading diagonal of $F(s)$ are polynomials of degree u ; but as a result of the restriction on the degree of the numerator terms of the transfer functions in the matrix $Q(s)$, discussed in §2 above, the non-diagonal elements are of degree $u-1$ at most. The coefficient of s^u in each of the diagonal elements of $F(s)$ can be taken as unity. There are u other coefficients of powers of s (including the constant term) in each of the diagonal elements and u coefficients in each of the non-diagonal elements. In calculating $F(s)$ from $\Phi(s)$, or equivalently from the K_r and $\lambda_r, r = 1, 2, \dots, m$, there are therefore nu coefficients to be determined in each row of $F(s)$. Since $|F(s)|$ is of degree nu we have $m = nu$. We obtain from (37) a set of nu non-homogeneous equations in the coefficients of each row of $F(s)$. It can be shown that the equations in each set are linearly independent. Thus all the coefficients in $F(s)$ are uniquely determined.

Given $F(s)$ we can find $Q(s)$ from

$$Q(s) = P^{-1}(s) F(s), \quad (56)$$

which follows from (21). It will be noticed that the elements of the diagonal matrix $P(s)$ are identical with the elements on the leading diagonal of $F(s)$. The numerator and denominator polynomials of the elements of $Q(s)$ as calculated from (56) may have common factors which can be cancelled.

Given $F(s)$ and the K_r and λ_r , $G(s) G'(-s)$ is readily obtained from the equation

$$G(s) G'(-s) = \sum_{r=1}^m \frac{F(s) K_r F'(-s)}{s - \lambda_r} + \sum_{r=1}^m \frac{F(s) K'_r F'(-s)}{-s - \lambda_r} \quad (57)$$

which is derived from (27) and (30). Using (19) and (22) we then obtain

$$\theta(s) = P^{-1}(s) G(s) G'(-s) P'^{-1}(-s). \quad (58)$$

The numerator and denominator polynomials of the elements of $\theta(s)$ as calculated from (58) may also have common factors which can be cancelled.

7. CALCULATION OF $Q(s)$ AND $\theta(s)$ FROM $\Phi(s)$ WHEN ALL THE EQUATIONS ARE NOT OF THE SAME ORDER, BUT THE DISTURBANCES ARE NOT CROSS-CORRELATED

If all the equations in a system are not of the same order they cannot all be identified unless further restrictions are placed on the system. Consider a system $F(D)x(t) = G(D)\zeta(t)$ in which the equations are not all of the same order, and in which the arrangement of the equations is such that the first equation is of the lowest order and the order of each successive equation is at least as high as that of the one preceding it. If (27) is satisfied by the matrices $F(s)$ and $G(s)$ of this system it is also satisfied by $\mu F(s)$ and $\mu G(s)$ where μ is an arbitrary matrix of constants. We can add to any row of $F(s)$ a multiple of an earlier row corresponding to an equation of lower order without violating the condition that the non-diagonal elements in each row be of lower degree than the diagonal elements; but we cannot add a multiple of any succeeding row without violating this condition. As a result of this condition, therefore, μ is restricted to a triangular matrix with zeros above the leading diagonal. Given $\Phi(s)$ for this system, the first equation is therefore determinate (and so are any other equations of the same order as the first one) but the remaining equations of higher order than the first one are not identified without further restrictions on the system.

If it is known that the disturbances to the system are not cross-correlated, $G(s) G'(-s)$ must be diagonal. If $\mu F(s)$ and $\mu G(s)$ are to be admissible solutions, therefore, μ must satisfy the further restriction that the non-diagonal elements of $\mu G(s) G'(-s) \mu'$ be identically zero. Writing g_{ii} for the diagonal elements of $G(s) G'(-s)$ the first part of the array formed by the elements lying above the leading diagonal of the matrix product $\mu G(s) G'(-s) \mu'$ is found to be

$$\begin{array}{lll} \mu_{11} g_{11} \mu_{21} & \mu_{11} g_{11} \mu_{31} & \mu_{11} g_{11} \mu_{41} \\ \mu_{21} g_{11} \mu_{31} + \mu_{22} g_{22} \mu_{32} & \mu_{21} g_{11} \mu_{41} + \mu_{22} g_{22} \mu_{42} & \\ & \mu_{31} g_{11} \mu_{41} + \mu_{32} g_{22} \mu_{42} + \mu_{33} g_{33} \mu_{43} & \end{array}$$

Since u_{ii} cannot be made zero, we see from the first element in the array that if $g_{11} \neq 0$, μ_{21} must vanish. From the second row of the array we see that if $g_{11} \neq 0$ and $g_{22} \neq 0$, μ_{31} and μ_{32} must vanish. Similarly, from the third row, μ_{41} , μ_{42} and μ_{43} must vanish if g_{11} , g_{22} and g_{33}

are non-zero. Continuing in this way it can be shown that all the elements μ_{ij} , ($i > j$) must vanish if all the elements g_{ii} , ($i = 1, 2, \dots, n-1$) are non-zero. The matrix μ is then diagonal, and all the equations in the system are identified.

To calculate $F(s)$ for the system we have just been considering, given $\Phi(s)$ we first note that since the equations of lowest order are identified without restriction of the disturbance covariance functions, it will be possible to find coefficients in at least one of the rows of $F(s)$ which satisfy the appropriate set of scalar equations derived from (37) for all the value of r . The number of equations will be greater than the number of non-zero coefficients in this row, but they will form a consistent set with a sufficient number of independent equations to determine the non-zero coefficients. In the present case the first equation in the system is of lowest order, so the coefficients in the first row of $F(s)$ can be found in this way.

From (27) we have
$$F(s) \Phi(s) F'(-s) = G(s) G'(-s). \quad (59)$$

On the assumption that $G(s) G'(-s)$ is diagonal the first column of the matrix product in (59) yields the equations

$$\sum_{j=1}^n \sum_{q=1}^n F_{ij}(s) \Phi_{jq}(s) F_{1q}(-s) = 0 \quad (i = 2, 3, \dots, n). \quad (60)$$

If we give $\Phi_{jq}(s)$ and $F_{1q}(-s)$ in (60) their known numerical values, multiply through by $p(s)p(-s)$ and equate the coefficients of each power of s we obtain for each value of i a consistent set of linear equations in the coefficients of the polynomials in the i th row of $F(s)$, which can be solved for these coefficients. $Q(s)$ can then be calculated from $F(s)$, and $\theta(s)$ is given by

$$\theta(s) = Q(s) \Phi(s) Q'(-s). \quad (61)$$

8. STOCHASTIC DIFFERENCE EQUATION FOR DISCRETE POINTS ON $x(t)$

Let $\Gamma(z)$ be the two-sided generating function of the discrete functions ϕ_τ defined in (39), so that

$$\Gamma(z) = \sum_{\tau=-\infty}^{\infty} \phi_\tau z^\tau. \quad (62)$$

Since $\phi_{-\tau} = \phi'_\tau$ we have from (40)

$$\Gamma(z) = M(z) + M'(z^{-1}) - \phi_0. \quad (63)$$

Substituting (44) in (63) we obtain

$$\Gamma(z) = U^{-1}(z) V(z) + V'(z^{-1}) U'^{-1}(z^{-1}) - \phi_0 \quad (64)$$

or

$$U(z) \Gamma(z) U'(z^{-1}) = W(z) W'(z^{-1}), \quad (65)$$

where $W(z)$ is any matrix which satisfies the equation

$$W(z) W'(z^{-1}) = V(z) U'(z^{-1}) + U(z) V'(z^{-1}) - U(z) \phi_0 U'(z^{-1}). \quad (66)$$

It follows from (65) that $\Gamma(z)$ is also the generating function of the covariance functions of discrete variables x_t which are generated by the system of difference equations

$$U(E^{-1}) x_t = W(E^{-1}) \epsilon_t \quad (67)$$

from stationary stochastic variables ϵ_t with covariance functions

$$E\{\epsilon_t \epsilon'_{t-\tau}\} = \begin{cases} I & \text{if } \tau = 0, \\ 0 & \text{if } \tau \neq 0. \end{cases}$$

(The symbol E^{-1} in (67) is the shift operator, defined by $E^{-h} y_t = y_{t-h}$.)

The discrete variables x_i in (67) may be identified with points at unit time intervals on the continuous variables $x(t)$ in (20), and it follows that the sampling properties of estimates of the continuous system (20) may be studied by considering the sampling properties of estimates of the discrete system (67). The work done by Bartlett (1946, 1955), M. G. Kendall (1945, 1946, 1949), Quenouille (1947, 1957), Whittle (1952, 1953), Wold (1949) and others on sampling properties and goodness-of-fit tests for autoregressive, moving average and mixed processes is therefore relevant to the estimation of parameters in systems of stochastic differential equations. Most of the theoretical results which have been obtained, however, are valid only for large samples.

Some sampling experiments are being carried out on systems of the type considered in this paper. The results will be published later.

REFERENCES

- BARTLETT, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time series. *J. Roy. Statist. Soc. B*, **8**, 27.
- BARTLETT, M. S. (1955). *Stochastic Processes*. Cambridge University Press.
- FRAZER, R. A., DUNCAN, W. J. & COLLAR, A. R. (1950). *Elementary Matrices*. Cambridge University Press.
- JAMES, H. M., NICHOLS, N. B. & PHILLIPS, R. S. (1947). *Theory of Servomechanisms*. New York: McGraw Hill.
- KENDALL, M. G. (1945). On the analysis of oscillatory time series. *J. Roy. Statist. Soc.* **108**, 93.
- KENDALL, M. G. (1946). *Contributions to the Study of Oscillatory Time Series*. Cambridge University Press.
- KENDALL, M. G. (1949). Tables of autoregressive series. *Biometrika*, **36**, 267.
- LANING, J. H. & BATTIN, R. H. (1956). *Random Processes in Automatic Control*. New York: McGraw Hill.
- MILLER, K. S. (1956). *Engineering Mathematics*. London: Constable and Co. Ltd.
- QUENOUILLE, M. H. (1947). A large-sample test for the goodness-of-fit of autoregressive schemes. *J. Roy. Statist. Soc. A*, **110**, 123.
- QUENOUILLE, M. H. (1957). *The Analysis of Multiple Time Series*. London: Charles Griffin and Co. Ltd.
- WHITTLE, P. (1952). Tests of fit in time series. *Biometrika*, **39**, 309.
- WHITTLE, P. (1953). The analysis of multiple stationary time-series. *J. Roy. Statist. Soc. B*, **15**, 125.
- WIDDER, D. V. (1946). *The Laplace Transform*. Princeton University Press.
- WOLD, H. (1949). A large-sample test for moving averages. *J. Roy. Statist. Soc.* **11**, 297.

DESIGN OF EXPERIMENTS IN NON-LINEAR SITUATIONS

BY G. E. P. BOX AND H. L. LUCAS

*Statistical Techniques Research Group, Princeton University
and Institute of Statistics, North Carolina State College*

1. OUTLINE OF THE PROBLEM

We suppose that some response η is a known function

$$\eta = f(\xi_1, \xi_2, \dots, \xi_k; \theta_1, \theta_2, \dots, \theta_p) = f(\xi; \theta) \quad (1)$$

of k variables whose levels are denoted by the elements $\xi_1, \dots, \xi_i, \dots, \xi_k$ of the vector ξ and of p parameters $\theta_1, \dots, \theta_r, \dots, \theta_p$ elements of the vector θ , and that this function is not necessarily linear in either the variables or the parameters. We further suppose that it is possible to perform a series of experiments or trials in each one of which the response is observed at a particular chosen set of levels of the variables decided in advance.

The problem here considered is that of selecting a programme of trials such that they may be expected to provide results from which the p parameters can be estimated with high accuracy. In general the experimental programme may be defined by an $N \times k$ matrix $D = \{\xi_{iu}\}$ called the design matrix. The u th row ξ'_u of this matrix with elements $\xi_{1u}, \dots, \xi_{iu}, \dots, \xi_{ku}$ provides the levels of the k variables at which the response is to be observed in the u th trial.

1.1. Example

The following example illustrates the situation we have in mind. Suppose a consecutive chemical reaction is under study in which a substance A decomposes to form substance B which then in turn decomposes to form substance C . With the assumptions that the reactions are first order and irreversible it may be shown that, after time ξ_1 has elapsed, the yield η of intermediate product B is given by

$$\eta = \frac{\theta_1}{\theta_1 - \theta_2} \{ \exp(-\theta_2 \xi_1) - \exp(-\theta_1 \xi_1) \}, \quad (2)$$

where θ_1 and θ_2 are constants measuring the specific rates of the first and second decompositions, respectively.

The problem is to choose a set of times $\xi_{1u}, u = 1, 2, \dots, N$, at which to observe the yield so that from these observations θ_1 and θ_2 can be estimated as accurately as possible, allowing for experimental error in the observations. In this example the design matrix D would simply consist of a single column whose N elements were the times $\xi_{11}, \dots, \xi_{1u}, \dots, \xi_{1N}$ at which the yield was to be observed and θ would contain the two elements θ_1 and θ_2 . More generally there will be k variables (for example, $\xi_1 = \text{time}$, $\xi_2 = \text{temperature}$, $\xi_3 = \text{concentration}$, etc.) and the design matrix will contain k columns.

1.2. Practical limits to levels of variables

In practice, levels at which variables such as time, temperature, concentration and dosage can be set are usually restricted. For example, in the study of the chemical reaction above, the feasible range for the variable time (ξ_1) will be from zero to some value $\xi_1(\text{max})$ which will be the longest time which could be contemplated in an actual experiment. In general, there

will exist some experimental region R in the ξ space which delimits the area within which trials can actually be conducted. In a chemical reaction, for example, such limitation may occur from explosive hazards associated with certain combinations of temperature and pressure. In other cases it may be simply that it is experimentally inconvenient to conduct trials at certain combinations of conditions. In some cases the experimental region can be defined by a series of inequalities in the ξ 's of the type

$$\xi_i(\min) \leq \xi_i \leq \xi_i(\max) \quad (i = 1, 2, \dots, k).$$

In general, however, the definition of R may be more complicated.

1.3. Criterion for selection of design

We denote the response observed at the u th set of experimental conditions by y_u and we suppose that

$$E(y_u) = \eta_u = f(\xi_u, \theta), \quad (3)$$

and
$$E(y_u - \eta_u)(y_v - \eta_v) = \begin{cases} \sigma^2 & u = v \\ 0 & u \neq v \end{cases} \quad (u, v = 1, 2, \dots, N), \quad (4)$$

where, in general, σ^2 is unknown. The true values of the parameters are denoted by $\theta_{10}, \theta_{20}, \dots, \theta_{p0}$, the elements of the vector θ_0 . The partial derivatives of the response function with respect to the r th parameter θ_r for the u th set of experimental conditions, taken at the point θ_0 , is denoted by

$$f_{ru} = \left[\frac{\partial f(\xi, \theta)}{\partial \theta_r} \right]_{\theta=\theta_0}. \quad (5)$$

Finally, the $N \times p$ matrix of these derivatives is

$$\mathbf{F} = \{f_{ru}\}. \quad (6)$$

It is well known that the least squares estimates $\hat{\theta}$ obtained by minimizing the sum of squares

$$\sum_{u=1}^N \{y_u - f(\xi_u, \theta)\}^2 \quad (7)$$

and given by solving the normal equations

$$\sum_{u=1}^N \{y_u - f(\xi_u, \hat{\theta})\} \{\partial f(\xi_u, \theta) / \partial \theta_r\}_{\theta=\hat{\theta}} = 0 \quad (r = 1, 2, \dots, p) \quad (8)$$

have a variance covariance matrix which is approximated by $(\mathbf{F}' \mathbf{F})^{-1} \sigma^2$.

We shall proceed by attempting to choose $\mathbf{D}$ so that the determinant $|(\mathbf{F}' \mathbf{F})^{-1}|$ is made as small as possible. If we assume that each $f(\xi_u, \theta)$ is approximately linear in the neighbourhood of θ_0 we are thus minimizing Wilk's generalized variance for the estimates. The logic of this choice, also made by Wald (1943), can be seen by noting that if the estimates are approximately normally distributed, the determinant $|(\mathbf{F}' \mathbf{F})^{-1}|$ is proportional to the volume contained within any specific ellipsoidal probability contour for θ about θ_0 in the space of the parameters. The criterion chosen ensures that any such probability contour includes the smallest volume.

To understand the significance of this criterion in general non-linear situations it seems desirable to consider the relationship between sample space, the 'solution locus' within sample space, and parameter space. The point θ in parameter space can be mapped on to the point $P(\theta)$ in sample space with co-ordinates $\eta_u = f(\xi_u, \theta)$ ($u = 1, \dots, N$). We define the part of sample space covered by this mapping as the 'solution locus'.

Now, given any point $P(\theta)$ on the solution locus, we can make an orthogonal transformation $\eta = H\zeta$ of co-ordinates in sample space such that the solution locus in the immediate neighbourhood of $P(\theta)$ coincides with part of the p -dimensional hyperplane $\zeta_i = \text{constant}$ ($i = p+1, \dots, N$). We can then interpret $|(F'F)^{-1}|$ as the square of the Jacobian of the transformation from co-ordinates $(\zeta_1, \dots, \zeta_p)$ in sample space to $(\theta_1, \dots, \theta_p)$ in parameter space, when $P(\theta) = P(\theta_0)$. So the criterion that $|(F'F)^{-1}|$ is to be made as small as possible means geometrically that the p -dimensional volume of that part of parameter space associated with a region of prescribed small volume round $P(\theta_0)$ on the solution locus shall be as small as possible. Since we do not know θ_0 in advance, we must assume that we can find a design that makes this Jacobian about as small as possible for all points $P(\theta)$ near $P(\theta_0)$.

Let us now assume that the part of the solution locus near $P(\theta_0)$ can be approximated by a p -dimensional linear subspace of sample space. If we also assume that the experimental error distributions are normally distributed as well as satisfying (3) and (4), then the confidence region for θ based on the likelihood ratio criterion consists of all points associated with points on the solution locus within a p -dimensional sphere centred on the least squares estimate $P(\hat{\theta})$, and with radius independent of the experimental design. Any design that gives an adequate approximation to the smallest possible value of the Jacobian for all points round $P(\theta_0)$ therefore approximately minimizes the p -dimensional volume of the confidence region in parameter space.

1.4. *Necessity for preliminary estimates*

We see that the efficiency of different possible designs depends upon the matrix F whose elements are the values of the derivatives of the response function with respect to the θ 's at $\theta = \theta_0$. Now the derivatives $\partial f / \partial \theta_r$ can only be independent of the values actually taken by the parameter θ_r if the response function f is linear in θ_r . For non-linear response functions the values of the derivatives, and hence the efficiency of any particular design, depend upon the actual values of the p -parameters. If we are to suppose that effective design is possible at all, we must also assume therefore that *something* is known about the values of the parameters in advance.

In practical problems it will almost invariably be the case that some such information is available, and this will then provide the basis of a first design. As is inevitably the case here and elsewhere, the real effectiveness of the design will depend upon the reliability of the information on which it is based. In this paper, therefore, we assume that preliminary values θ^* of the parameters are available and will proceed so that the design would be optimal if these quantities were the true values. Some further discussion of the general situation for which this theory is appropriate is given by Box (1957).

1.5. *Possible extensions of the problem*

It is clear that in some examples normality of distribution and constancy of variance would not be possible at all points on the response function if only because the variation might be limited either upwards or downwards. Two obvious extensions, therefore, of this work would be to generalize to a maximum likelihood solution in which any distribution is considered or to consider weighted least squares. Valuable work in this direction has already been carried out by Elfving (1952) and Chernoff (1953). Again, in some cases we might wish to make a transformation of the model before the present theory was applied.

Apart from questions of appropriate distribution assumptions it should be possible to extend this work in at least two other directions. In the first place in different problems the reliability of the preliminary values θ^* will be different, and we might take account of this by using some rough estimate of the prior probability distribution. Again, some observations might be cheaper than others. One could then attempt to find the design giving the most information for a given cost. Finally, it would be desirable in many circumstances to use a sequential procedure in which estimates from a preliminary experiment would be used to plan a second experiment, the results of which would in turn be used to plan a third experiment, and so on. These possibilities will not be further discussed here but we hope it may be possible to pursue them in later work.

2. SOLUTION

Suppose that we are given the functional form

$$\eta_u = f(\xi_u, \theta) \quad (9)$$

and a set of preliminary values θ^* . Then we shall choose the design D such that $|F^*F^*|$ is a maximum within the experimental region R in the ξ -space, where

$$F^* = \{f_{ru}^*\} \quad (10)$$

and

$$f_{ru}^* = \left[\frac{\partial f(\xi_u, \theta)}{\partial \theta_r} \right]_{\theta=\theta^*} \quad (11)$$

To limit to manageable proportions the possibilities we need to consider, we will in this paper assume that the number of trials (that is the number of combinations of levels of the variables) is equal to the number p of parameters to be estimated. Each of the p trials could of course be replicated to attain any desired degree of precision, but it should be noted that an r -fold replication of a design that minimizes the generalized variance using p observations, does not necessarily minimize the generalized variance for all possible experiments involving rp observations. Experiments in which the number of parameters estimated is equal to the number of trials are not perhaps intuitively appealing. This is (i) because one is usually not completely assured that the model is correct; so that one ordinarily has in fact a twofold objective: to check the model and to estimate the parameters in it if it proves to be representationally adequate, and (ii) because designs in which there are more than the minimum number of points may provide greater protection against gross errors in guessing θ^* .

In order for a design to be efficient in 'checking' the model, we must be specific about what sorts of departure from the model we suspect. In so far as this can be done, we should be able to write an 'alternative' wider model which makes allowance for these possible departures *in terms of extra parameters to be estimated*. In this way, within the framework discussed we can achieve both objectives mentioned above. As a simple illustration consider the chemical reaction discussed earlier. Here A was decomposed to form B which in turn decomposed to C . In a particular problem the second decomposition might be unlikely, in which case the primary model in mind for the production of B would be

$$\eta = 1 - e^{-\theta_1 \xi} \quad (12)$$

However, in the light of possible decomposition of B to C it would be appropriate to plan the experiment so as to be effective in estimating both θ_1 and θ_2 in the model of equation (2).

When nothing is known about the type of departures which may occur we are confronted once more with the fact already mentioned that the effectiveness of a design must depend on

the reliability of information on which it is based. If this information is non-existent then we have no basis on which to plan. In practice, however, the situation may arise where some information is available which is difficult to express in terms of extra parameters in the model, but is manifested in feelings about the way the response surface may be distorted. In this case further points might be added to the design in an intuitive way, so as to tie down doubtful areas of the surface, while at the same time bearing in mind the locations already determined which ensure maximum reliability for the principal parameters when the model is adequate.

When gross errors may occur in guessing θ^* the present theory can be used to explore the changes that would occur in the position of the optimal p design points with changes in θ^* over the region of uncertainty and the losses in efficiency which occurred with given errors in θ^* . If more than p -points were available these might then be chosen so as to minimize the loss of efficiency as far as possible.

2.2. A geometrical formulation

For that class of designs for which the number of trials is equal to the number of parameters, $\mathbf{F}^*$ is a square ($p \times p$) matrix. Consequently,

$$|\mathbf{F}^{*'}\mathbf{F}^*| = |\mathbf{F}^*|^2 \quad (13)$$

and we should choose the design $\mathbf{D}$ so that the modulus of $|\mathbf{F}^*|$ is a maximum in the experimental region R .

In understanding the problem to be solved and in showing the properties of solutions in simple cases, the following geometrical formulation has been found of help. Consider a p -dimensional space in which the co-ordinates are the derivatives

$$\left(\frac{\partial f(\xi, \theta)}{\partial \theta_r} \right)_{\theta=\theta^*} = \phi_r(\xi) \quad (r = 1, 2, \dots, p), \quad (14)$$

where, it will be noted, $\phi_r(\xi)$ is a function of ξ only. The surface in the space of the derivatives generated by all permissible values of ξ we call the 'design locus'. Thus with any particular choice of p 'design points', $\xi_1, \xi_2, \dots, \xi_u, \dots, \xi_p$, in the ξ -space for a particular design matrix $\mathbf{D}$, there will be corresponding points on the design locus in the space of the derivatives $\phi_r(\xi)$.

Now the volume of a simplex in p -space formed by p -points

$$(z_{11}, z_{12}, \dots, z_{1p}); \dots; (z_{s1}, z_{s2}, \dots, z_{sp}); \dots; (z_{p1}, z_{p2}, \dots, z_{pp})$$

and the origin $(0, 0, \dots, 0)$ is proportional to the determinant of the $p \times p$ matrix $\{z_{sr}\}$. Consequently, we must find p design points $\xi_1, \dots, \xi_u, \dots, \xi_p$ for which the corresponding points in the space of the derivatives, with co-ordinates defined by equation (14), form with the origin the simplex of greatest volume.

2.3. Example

As a simple illustration, consider the example given in § 1.1 of a consecutive first-order chemical reaction. Suppose that we are given preliminary guesses θ_1^*, θ_2^* for the values of the parameters. We wish to choose two values ξ_{11}, ξ_{12} of time so that the best estimates will be available for θ_1, θ_2 . We have

$$\eta = f(\xi_1; \theta_1, \theta_2) = \frac{\theta_1}{\theta_1 - \theta_2} \{ \exp(-\theta_2 \xi_1) - \exp(-\theta_1 \xi_1) \}. \quad (15)$$

2.5. Numerical solution

Although in particular examples there is an analytic solution to the problem of maximizing the modulus of the determinant $|F^*|$, often numerical methods must be employed. Such methods are attractive because they allow us to obtain solutions even when the function $f(\xi, \theta)$ is not given explicitly. In chemical examples, for instance, a kinetic mechanism can often be postulated in terms of a series of differential equations which may have no explicit solution. The equations may be integrated numerically, however, on a computer and the quantities $\phi_r(\xi)$ for any trial value ξ obtained from difference formulae. Iterative methods may then be employed to find the maximum value of the modulus of the determinant.

In the particular example studied it is seen from Fig. 3 that in the neighbourhood of the optimum solution the surface is well conditioned, containing no oblique ridges, and that consequently a numerical method in which the variables were adjusted one at a time would be successful in determining the maximum. It is fairly easy to see that the satisfactory state of conditioning indicated by the 'symmetrical' maxima found in this problem cannot be expected to be general. Often the Δ surface would contain attenuated ridges which in the later stages of iteration would give rise to difficulties in numerical solution when the method of adjusting 'one variable at a time' or 'steepest ascent' procedures were used. A method which can be employed when the final solution is approached is the second-order procedure similar to that employed by Koshal (1933) in solving maximum likelihood equations. This consists of fitting a second-degree polynomial to values of the determinant calculated at suitably arranged series of points.

Calculations for improved values of ξ_{11} and ξ_{12} , starting from the values $\xi_{11} = 1.10$, $\xi_{12} = 6.50$ (obtained from Fig. 2) are shown for the present example in Table 1.

Table 1. Calculation of improved values of maximum

ξ_{11}	ξ_{12}	$-\Delta$
1.10	6.50	0.803 336
1.20	6.50	0.807 992
1.30	6.50	0.806 709
1.10	7.00	0.804 758
1.10	7.50	0.798 920
1.20	7.00	0.809 632

Writing $x_1 = 5(\xi_{11} - 1.10)$, $x_2 = (\xi_{12} - 6.50)$, the second-degree interpolation polynomial is

$$-\Delta = 0.803336 + 0.015251x_1 + 0.010104x_2 - 0.011878x_1^2 - 0.014250x_2^2 + 0.000872x_1x_2$$

which has a maximum at

$$x_1 = 0.66, \quad x_2 = 0.37,$$

that is

$$\xi_{11} = 1.23, \quad \xi_{12} = 6.87.$$

A second iteration provided values in substantial agreement with these, namely, $\xi_{11} = 1.23$ and $\xi_{12} = 6.86$. It is clear that because of difficulties of possible multiple maxima and the complexities of the Δ surface, considerable caution is necessary in applying numerical

methods to problems in which there are a large number of parameters. It is hoped to discuss such cases in a later paper. The examples which follow contain only a few parameters but nevertheless are thought to be of some inherent interest.

3. EXAMPLES OF NON-LINEAR DESIGNS

Consider the model

3.1. Simple decay and growth law

$$\eta = \theta_1 \exp(\theta_2 \xi_1), \quad 0 \leq \xi_1(\min) \leq \xi_{11} < \xi_{12} \leq \xi_1(\max). \quad (21)$$

With ξ representing time and with $\theta_1 > 0$, $\theta_2 < 0$, this is the well-known first-order decay law. With $\theta_1 > 0$, $\theta_2 > 0$, the function can represent growth phenomena. In application both lower and upper limits $\xi_1(\min)$ and $\xi_1(\max)$ may be imposed corresponding to the earliest and latest points in time at which observations can be made. The value of the determinant is

$$\Delta = \theta_1 \exp\{\theta_2(\xi_{11} + \xi_{12})\}(\xi_{12} - \xi_{11}). \quad (22)$$

The modulus of which for $\theta_2^* < 0$ attains its maximum value when

$$\left. \begin{aligned} \xi_{11} &= \xi_1(\min), \\ \xi_{12} &= \xi_1(\min) - 1/\theta_2^*. \end{aligned} \right\} \quad (23)$$

Substituting these values into equation (21) we see that the best design will be that in which a proportion $e^{-1} = 36.8\%$ of the amount of substance present at the first observation remains at the second observation.

For $\theta_2^* > 0$ the modulus of Δ is a maximum when

$$\left. \begin{aligned} \xi_{11} &= \xi_1(\max) - 1/\theta_2^*, \\ \xi_{12} &= \xi_1(\max). \end{aligned} \right\} \quad (24)$$

The first observation should thus be made at a time when 36.8% of the growth expected at the final observation has occurred.

If instead of the amount of substance decayed we consider the amount of product formed from a simple decay law, we have

$$\eta = \theta_1 \{1 - \exp(\theta_2 \xi_1)\} \quad (25)$$

for $0 \leq \xi_1(\min) \leq \xi_{11} < \xi_{12} \leq \xi_1(\max)$.

The value of the determinant is

$$\Delta = \theta_1 [\{1 - \exp(\theta_2 \xi_{12})\} \xi_{11} \exp(\theta_2 \xi_{11}) - \{1 - \exp(\theta_2 \xi_{11})\} \xi_{12} \exp(\theta_2 \xi_{12})], \quad (26)$$

the modulus of which has a maximum when

$$\xi_{12} = \xi_1(\max), \quad (28)$$

$$\xi_{11} = -\frac{1}{\theta_2^*} - \frac{\xi_{12} \exp(\theta_2^* \xi_{12})}{1 - \exp(\theta_2^* \xi_{12})}.$$

Substituting these values into equation (25) and letting γ_2 be the fraction by which the yield at time $\xi_{12} = \xi_1(\max)$ falls short of the maximum obtainable, we find that the optimum value of ξ_{11} corresponds to a yield falling short of the maximum by a fraction

$$\gamma_1 = \exp\{-[1 + \gamma_2 \log \gamma_2 / (1 - \gamma_2)]\}. \quad (29)$$

In particular if the second observation is taken after a sufficiently long time so that the asymptotic value is almost achieved, then γ_2 is close to zero and $\gamma_1 = 1/e$, so that the first observation should be made when the process is 63.2 % complete. As a second example, if the second observation is taken at a time when the process is 50 % complete, then $\gamma_1 = 0.73$ and the first observation should be made at a time when the process is 27 % complete.

3.2. Mitscherlich equation

Consider the model

$$\eta = \theta_1 + \theta_2 \exp(\theta_3 \xi_1) \quad (30)$$

for

$$0 \leq \xi_1(\min) \leq \xi_{11} < \xi_{12} < \xi_{13} \leq \xi_1(\max).$$

With $\theta_3 < 0$ and $\theta_2 < 0$ this expression represents the well-known Mitscherlich law of diminishing returns. For example, in crop experiments with ξ_1 the amount of added fertilizer and η the amount of growth, $\theta_1 + \theta_2$ represents the growth when no fertilizer is added and θ_1 the hypothetical growth obtained from an infinite amount of fertilizer; θ_3 measures the rate at which the response to each additional increment of fertilizer decreases. The same model has been employed to relate the weight η of chickens to their cumulative feed consumption ξ_1 (Hendricks *et al.* 1932). With $\theta_3 > 0$ this model has been used to represent early growth (Brant, 1951).

The value of the determinant is

$$\Delta = \theta_2 [\xi_{11} \exp(\theta_3 \xi_{11}) \{ \exp(\theta_3 \xi_{13}) - \exp(\theta_3 \xi_{12}) \} + \xi_{12} \exp(\theta_3 \xi_{12}) \times \{ \exp(\theta_3 \xi_{11}) - \exp(\theta_3 \xi_{13}) \} + \xi_{13} \exp(\theta_3 \xi_{13}) \{ \exp(\theta_3 \xi_{12}) - \exp(\theta_3 \xi_{11}) \}]. \quad (31)$$

It is readily shown that irrespective of the signs of $\theta_1, \theta_2, \theta_3$, the optimum design is obtained with

$$\xi_{11} = \xi_1(\min), \quad (32)$$

$$\xi_{12} = -\frac{1}{\theta_3^*} + \frac{\xi_{11} \exp(\theta_3^* \xi_{11}) - \xi_{13} \exp(\theta_3^* \xi_{13})}{\exp(\theta_3^* \xi_{11}) - \exp(\theta_3^* \xi_{13})}, \quad (33)$$

$$\xi_{13} = \xi_1(\max). \quad (34)$$

In particular in a fertilizer experiment if $\xi_{11} = \xi_1(\min)$ corresponded to no dressing of fertilizer and $\xi_{13} = \xi_1(\max)$ corresponded to a dressing for which the response fell short of the maximum by a certain fraction γ_2 , then ξ_{12} would be a dressing for which the response fell short of the maximum possible by a fraction γ_1 where again γ_1 would be given by (29). When $\xi_1(\max)$ is sufficiently large to give an almost maximal response then the intermediate value of the design should be that expected to yield a response of 63.2 % of the maximum possible.

3.3. Simple first-order decay with rate a function of temperature

Consider the model

$$\eta = \exp(-\theta_1 \xi_1 e^{-\theta_2 \xi_2}) \quad (35)$$

for

$$0 \leq \xi_1(\min) \leq \xi_{11} < \xi_{12} \leq \xi_1(\max),$$

$$\xi_2(\min) \leq \xi_{21} < \xi_{22} \leq \xi_2(\max).$$

This would apply for example where η represented the fractional amount remaining at time $t = \xi_1$ of a material decaying according to the first-order law, when the rate was temperature-

sensitive. Assuming that the temperature effect follows the Arrhenius rule, θ_1 is the specific rate at some absolute temperature T_0 , θ_2 is proportional to the 'activation energy' and $\xi_2 = T^{-1} - T_0^{-1}$ is the deviation of the reciprocal of the temperature, T , from the reciprocal of the base temperature T_0 .

We find for preliminary values θ_1^* , θ_2^* that

$$\begin{aligned}\phi_1(\xi_1, \xi_2) &= -\xi_1 \exp\{-(\theta_2^* \xi_2 + \theta_1^* \xi_1 e^{-\theta_2^* \xi_2})\} = \frac{\eta^* \ln \eta^*}{\theta_1^*}, \\ \phi_2(\xi_1, \xi_2) &= -\theta_1^* \xi_2 \phi_1(\xi_1, \xi_2) = -\xi_2 \eta^* \ln \eta^*,\end{aligned}\quad (36)$$

where $\eta_u^* = \exp(-\theta_1^* \xi_{1u} e^{-\theta_2^* \xi_{2u}})$, the expected fraction of decaying substance remaining in the u th experiment. Thus

$$\text{mod } \Delta = \text{mod} \begin{vmatrix} \frac{\eta_1^*}{\theta_1^*} \ln \eta_1^* & -\xi_{21} \eta_1^* \ln \eta_1^* \\ \frac{\eta_2^*}{\theta_1^*} \ln \eta_2^* & -\xi_{22} \eta_2^* \ln \eta_2^* \end{vmatrix} = \frac{\eta_1^* \ln \eta_1^* \eta_2^* \ln \eta_2^*}{\theta_1^*} (\xi_{22} - \xi_{21}). \quad (37)$$

Now $-\eta \log \eta$ has its maximum value when $\eta = 1/e$. Also, if the range of permissible values for the times, ξ_1 , is sufficiently wide then the expression (37) is maximized by ensuring that $\xi_{22} - \xi_{21}$ shall achieve the maximum possible range $\xi_2(\text{max}) - \xi_2(\text{min})$ and adjusting the times of reaction so that the anticipated fractions remaining, η_1^* , η_2^* , are each equal to $1/e$. The actual values of ξ_{11} and ξ_{12} required in this case are obtained by substituting $\eta = e^{-1}$ in equation (35). The appropriate times ξ_{11} , ξ_{12} , associated with $\xi_2(\text{min})$ and $\xi_2(\text{max})$ are then given by

$$\begin{aligned}\xi_{11} &= \frac{1}{\theta_1^*} \exp\{\theta_2^* \xi_2(\text{min})\}, \\ \xi_{12} &= \frac{1}{\theta_1^*} \exp\{\theta_2^* \xi_2(\text{max})\}.\end{aligned}\quad (38)$$

It is instructive to consider a numerical example. Suppose the absolute temperature is restricted to the range 380–420° A. It is then convenient to take as an arbitrary base temperature the value $T_0 = 400^\circ \text{A}$. Suppose with time measured in minutes a guessed value for the rate θ_1 at this arbitrary base temperature is $\theta_1^* = 1.0$ and a guessed value for $\theta_2^* = 16,000$, then we have

$$\begin{aligned}\theta_1^* &= 1.0, \quad \theta_2^* = 16,000 \\ 380^\circ \text{A.} &\leq T = (\xi_2 + 0.0025)^{-1} \leq 420^\circ \text{A.}\end{aligned}$$

The optimum design will consist of two experiments, the first of which is performed at the minimum temperature of 380° A for time $\xi_{11} = 8.21$ min. and the second at the maximum temperature of 420° A for 0.15 min. These times are found by substituting in equation (38) and are simply those values at which, if the guessed values are correct and there were no experimental error, the fraction of product remaining would be $1/e$. The design is illustrated geometrically in Fig. 4. In this figure the design locus in the plane of $\phi_1(\xi_1, \xi_2)$, $\phi_2(\xi_1, \xi_2)$ is bounded by the values $\phi_1(\xi_1, \xi_2) = 0$ and $\phi_1(\xi_1, \xi_2) = -1/\theta_1^* e$, the largest and the smallest values, respectively, that this derivative can take. The straight lines in Fig. 4 are for constant ξ_2 (i.e. for constant temperature) and the curved lines for constant ξ_1 (i.e. for constant time). Where, as we have assumed above, the permissible range of ξ_2 dictates the

design points selected, the design locus is restricted to that portion of the plane within the triangle bounded by the three straight lines

$$\begin{aligned}\phi_1(\xi_1, \xi_2) &= -1/\theta_1^* e, \\ \phi_2(\xi_1, \xi_2) &= -\theta_1^* \phi_1(\xi_1, \xi_2) \xi_2(\min), \\ \phi_2(\xi_1, \xi_2) &= -\theta_1^* \phi_1(\xi_1, \xi_2) \xi_2(\max).\end{aligned}$$

The triangle so delineated is itself the triangle of largest area which can be formed with two points on the design locus and the origin with only temperature restricted.

The appropriate design for the case considered is indicated by the two heavy dots in the diagram. It will be seen from the figure that, if ξ_1 were the restricting variable, then a somewhat different result would be obtained, since now the triangle of maximum area would be achieved when the value of $\phi_1(\xi_1, \xi_2)$ fell somewhat short of $-1/\theta_1^* e$.

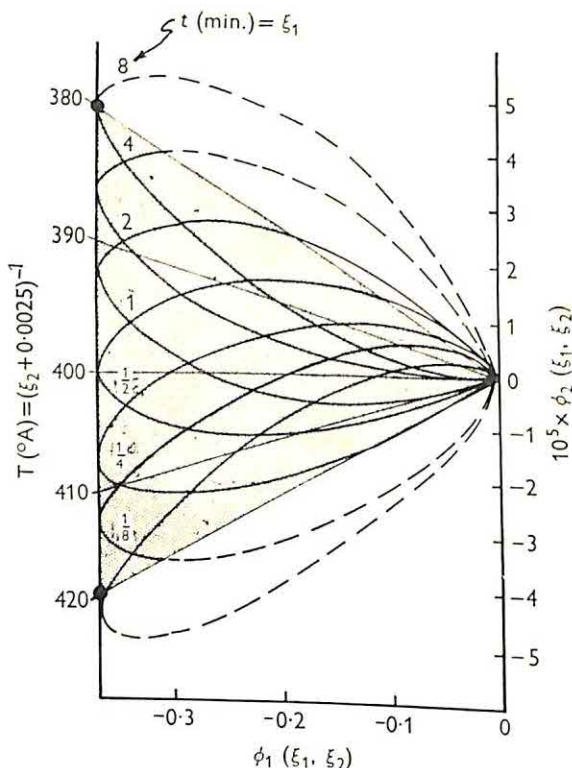


Fig. 4. The design locus for $\eta = \exp\{-1.0te^{-16,000(T^{-1}-0.0025)}\}$. The optimal solution for $T(\min) = 380^\circ\text{A.}$, $T(\max) = 420^\circ\text{A.}$ and $0 \leq \xi_1 < \infty$ is indicated by heavy dots.

4. CALCULATION OF THE ADJUSTMENTS

Suppose a design of this type has been obtained for assumed values of the parameters θ^* , and suppose that the N calculated values at these sets of conditions for $\theta = \theta^*$ are given by the column vector η^* . Suppose finally that observed responses from the N experiments performed are provided by the column vector y . Then proceeding in the usual way a $p \times 1$ vector of adjustments δ to be applied to the preliminary values θ^* are given to a first approximation by

$$\delta = (F^{*'}F^*)^{-1}F^{*'}(y - \eta^*)$$

or in the special case where the number of observations N is equal to the number of parameters p by

$$\delta = F^{*-1}(y - \eta^*).$$

In the particular example of § 3.3,

$$\mathbf{F}^* = \frac{1}{e} \begin{bmatrix} -1/\theta_1^* & \xi_{21} \\ -1/\theta_1^* & \xi_{22} \end{bmatrix}$$

and

$$\mathbf{y} - \boldsymbol{\eta}^* = \begin{bmatrix} y_2 - e^{-1} \\ y_1 - e^{-1} \end{bmatrix},$$

so that the adjusted values will be

$$\theta_1' = \theta_1^* + \delta_1,$$

$$\theta_2' = \theta_2^* + \delta_2,$$

where δ_1 and δ_2 are the 'regression coefficients' of the elements in $\mathbf{y} - \boldsymbol{\eta}^*$ on the columns of the matrix $\mathbf{F}^*$.

To see the meaning of this suppose that the value of θ_1^* were taken at a base temperature such that $\xi_{21} = -\xi_{22}$, then the adjustments in θ_1^* and θ_2^* are simply

$$\delta_1 = -(\bar{y} - e^{-1}) e \theta_1^*,$$

$$\delta_2 = (y_1 - y_2) e / 2 \xi_{21}.$$

As would be expected, therefore, when the observed values of y were equal but different from the value e^{-1} , this would imply that adjustment was needed only in the basic rate constant θ_1^* . Unequal values for y_1 and y_2 on the other hand would imply the necessity for adjustment in θ_2^* .

5. DISCUSSION

When parameters are jointly estimated from observations made at particular sets of conditions, it frequently happens that the estimates are highly correlated and consequently have very large variances. This may be simply the result of poor design in which case the first group of estimates may be used as starting values to design a more informative set of trials from which greatly improved estimates can be obtained. Not infrequently, however, it turns out that the unsatisfactory nature of the estimates is a property of the model itself, and that even the best design still yields highly correlated estimates. This point is well illustrated by models such as

$$\eta = \frac{1 + \theta_1 \xi_1}{\theta_2 + \theta_3 \xi_1}.$$

In accordance with the method already outlined, the optimal design will be found by choosing three points on the design locus which together with the origin form the tetrahedron of greatest volume. But the design locus for such a model often turns out to be a nearly straight line, so that for any reasonable range of the variable, ξ_1 , the volume of the tetrahedron and hence the value of the determinant is small. This arises from the nature of the model itself. Whatever values of the ξ 's are chosen in such examples, when one has found one set of values of the θ 's that most closely agrees with a given set of data, one can, by making suitable compensating changes in the parameters, obtain other quite different values of the θ 's that agree almost as well.

In this connexion it is important to distinguish between the problem we are considering here, which is that of finding designs most efficient for estimating the parameters $\boldsymbol{\theta}$, and a somewhat different problem which is *not* discussed here, that of finding designs such that the response $\eta = f(\boldsymbol{\xi}; \boldsymbol{\theta})$ is determined most precisely in a given region. A particular design may be such that the estimates, $\boldsymbol{\theta}$, of the parameters are highly correlated with large

variances and yet the response, η , is estimated reasonably accurately. The reason for this is not difficult to see. Instability in the θ 's which is *compensating* is by its very nature not such as will cause instability in the estimate of η .

The authors wish to acknowledge their indebtedness to a referee and to E. M. L. Beale for valuable suggestions for improving the presentation.

REFERENCES

- BOX, G. E. P. (1957). Use of statistical methods in the elucidation of basic mechanisms. *Bull. Inter. Statist. Inst.* **36**, part 3, 215-25.
- BRANT, J. W. A. (1951). Rate of early growth in domestic fowl. *Poult. Sci.* **30**, 343-61.
- CHERNOFF, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.* **24**, 586-602.
- ELFVING, G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Statist.* **23**, 255-62.
- HENDRICKS, W. A., JULL, M. A. & TITUS, H. W. (1932). The utilization of feed by chickens. I. The requirements for growth. *Poult. Sci.* **2**, 74-7.
- KOSHAL, R. S. (1933). Application of the method of maximum likelihood to the improvement of curves fitted by the method of least squares. *J. Roy. Statist. Soc.* **96**, 303-13.
- WALD, A. (1943). On the efficient design of statistical investigations. *Ann. Math. Statist.* **14**, 134-40.

THE ESTIMATION OF MISSING AND MIXED-UP OBSERVATIONS IN SEVERAL EXPERIMENTAL DESIGNS

BY J. D. BIGGERS

Department of Physiology, Royal Veterinary College, London

I. INTRODUCTION

The simple analysis of most experimental designs depends upon their fully balanced arrangement. Often observations are lost for a few of the units in the course of an experiment, and to overcome the consequent difficulties in the statistical analysis, because of non-orthogonality, the technique of inserting 'missing values' was introduced (Allan & Wishart, 1930; Yates, 1933). The values which are inserted minimize the error sum of squares.

In this paper the problem of calculating these missing values will be investigated from a general point of view for the following designs:

- (1) randomized block,
- (2) randomized block design with replication within units,
- (3) cross-over,
- (4) latin square and higher squares,
- (5) split-plot.

Adjustment of the variances of treatment comparisons is sometimes necessary after estimating missing values, but although this can be done by an extension of the present procedure it will not be discussed in this paper.

NOTATION

Scalars—letters in italics.

Vectors—lower case letters in Clarendon type.

Matrices—capital letters in Clarendon type.

Transposition of matrices and vectors is indicated by a prime.

Summation—two summation symbols will be used. If we consider the sequence x_i ($i = 1, 2, \dots, n$), we define ordinary summation over all values as $\Sigma_i x_i$. Also we define

$$\Sigma_{(i)} x_i = x_1 + x_3 + x_{n-1}, \text{ say,}$$

to mean summation over some subset of the subscript values—typically one relating to the missing values.

Associated observations—an observation in a design is denoted by $X_{ijk\dots}$, where the number of subscripts indicates the order of classification involved. Thus an observation in a randomized block requires two subscripts for its specification, and an observation in a Graeco-latin square requires four.

If we consider two observations they may have certain subscripts in common. This property enables us to describe the relation between pairs of observations; observations related in a specific way will be called *associates*. For example, we consider two observations

X_{fgh} and X_{uvw} . The following classes of associate can be recognized:

- (1) zero-order associates: $f \neq u, g \neq v, h \neq w$;

- (2) 1st-order associates: (i) i associates $f = u, g \neq v, h \neq w$;
 (ii) j associates $f \neq u, g = v, h \neq w$;
 (iii) k associates $f \neq u, g \neq v, h = w$;
- (3) 2nd-order associates: (i) ij associates $f = u, g = v, h \neq w$;
 (ii) ik associates $f = u, g \neq v, h = w$;
 (iii) jk associates $f \neq u, g = v, h = w$.

Particular associates may or may not exist depending on the experimental design.

A *potential observation* in a design will be indicated by a capital letter: $X_{ij...}$; if this is in fact a *missing observation* it will be indicated in lower case: $x_{ij...}$.

General solution

Let there be p independent missing values denoted by $x_i (i = 1, 2, \dots, p)$. The error sum of squares is a quadratic function of these values, $F(x_1, x_2, \dots, x_p)$. The vector of partial derivatives of this function is given by

$$2(\mathbf{x}'\mathbf{A} - \mathbf{q}')/N, \quad (1)$$

where $\mathbf{x}$ is a column vector of the p missing values,

$\mathbf{A}$ is a $p \times p$ symmetric matrix whose elements are determined by the experimental design and the distribution of the missing values,

$\mathbf{q}$ is a column vector whose elements are calculated from the available observations,

N is a constant determined by the experimental design.

The values of the x_i where $F(\mathbf{x})$ is a minimum are given by the solution of p simultaneous equations obtained by equating each of its partial derivatives to zero. These are, in matrix notation:

$$\mathbf{Ax} = \mathbf{q}. \quad (2)$$

Several methods may be used to solve equation (2). Most textbooks on the design and analysis of experiments give the formula for one missing observation in randomized block, cross-over and latin-square designs, and the special formulae for two or three missing observations have also been published for the randomized block design (Baten, 1939, 1952; Federer, 1951, 1955) and the latin-square design (Federer, 1955). The split-plot design has been dealt with for single missing observations by Anderson (1946) and Khargonkar (1948). Usually, where more than one missing observation exists in a design, it is recommended that the estimates be obtained by an iterative process based on the formula for one missing observation. Other methods are the covariance method (Bartlett, 1937; Coons, 1957) and the method of diagonal iteration (Healy & Westmacott, 1956) for use with automatic computers. None of these methods entails the explicit determination of the simultaneous equations. In this paper rules will be given which enable (2) to be written down rapidly for the different experimental designs. The main advantage of this procedure is that it enables the nature of the equations to be examined and this leads to simplifications in several important complex cases.

The matrix method has been the subject of a paper by Thompson (1956) on missing values in the randomized block design. This worker, however, has only dealt with a restricted distribution of missing values, presumably because values of $\mathbf{A}^{-1}$ are written down instead of $\mathbf{A}$. Very recently Wilkinson (1957) has shown how the matrix $\mathbf{A}$ can be used to obtain exact variances of treatment comparisons, and has also extended the method for the estimation of missing observations in the analysis of covariance.

II. RANDOMIZED BLOCK DESIGN

Suppose we have r blocks each containing t plots to which t treatments are allotted at random. We denote the observation made on the j th treatment in the i th block by X_{ij} ($i = 1, 2, \dots, r; j = 1, 2, \dots, t$). At the end of the experiment let p of these values be missing. It should be noted that the missing observations are indicated by only some values of the subscripts.

Let the i th block and the j th treatment contain c_i and d_j missing observations respectively. Then

$$p = \sum_i c_i = \sum_j d_j.$$

Missing observations in the same block are i or *block associates* and missing observations receiving the same treatment are called j or *treatment associates*. The missing observation x_{gh} will have therefore $(c_g - 1)$ block associates, $(d_h - 1)$ treatment associates and $\{p - (c_g + d_h - 1)\}$ zero associates.

Let B_i = total of the available observations in the i th block,

T_j = total of the available observations receiving the j th treatment,

G = total of all available observations.

The error term in the analysis of variance is given by

$$\sum_{(i)} \sum_{(j)} x_{ij}^2 - \sum_{(i)} (P_i + \sum_{(j)} x_{ij})^2/r - \sum_{(j)} (T_j + \sum_{(i)} x_{ij})^2/r + (G + \sum_{(i)} \sum_{(j)} x_{ij})^2/rt + \text{constant}. \quad (3)$$

Partial differentiation of (3) with respect to a particular missing observation, x_{gh} , and equating the derivative to zero, gives the equation

$$rtx_{gh} - r\Sigma_{(j)}x_{gj} - t\Sigma_{(i)}x_{ih} + \Sigma_{(i)}\Sigma_{(j)}x_{ij} = rB_g + tT_h - G.$$

Separation of terms corresponding to zero, i and j associates gives

$$(r-1)(t-1)x_{gh} + \overbrace{(1-r)\sum_{j \neq h}^{(i)} x_{gj}}^{i \text{ associates}} + \overbrace{(1-t)\sum_{i \neq g}^{(j)} x_{hi}}^{j \text{ associates}} + \overbrace{\sum_{i \neq g}^{(j)} \sum_{j \neq h}^{(i)} x_{ij}}^{\text{zero associates}} = rB_g + tT_h - G. \quad (4)$$

Expression (4) enables us to write down the coefficients of the x_{ij} , the B_g and T_h , and hence the matrix $\mathbf{A}$ and the vector $\mathbf{q}$ in (2). The rules which determine the elements of $\mathbf{A}$ are:

- (i) $a_{ij} =$ the coefficient of x_{ij} in the expression under consideration,

- (i) $(r-1)(t-1)$ for the missing observation under consideration,
(ii) $(1-r)$ for the i or block associates,
(iii) $(1-t)$ for the j or treatment associates,
(iv) 1 for the zero associates.
- In all practical situations the matrix $\mathbf{A}$ will be non-singular, and thus the equation (2) will have a unique solution.
- Example 1.* In a randomized block experiment ($r = 8; t = 5$). No

In all practical situations the matrix A will be non-singular, and thus the equation (2) will have a unique solution.

Example 1. Table 1 shows the results of a randomized block experiment ($r = 8; t = 3$). No missing values occurred in this experiment, but four have been omitted to provide the example. The missing values are calculated as follows:

- (1) With $t = 3$, the rows correspond to separate blocks. This is

- (1) Write out the table of results so that the rows correspond to the associates.
- (2) Obtain the coefficients of the associates. In the example these are 28 for the missing observation under consideration, -7 for the i associates, -4 for the j associates and 1 for the zero associates.
- (3) Insert in the table of results the letter x_{ij} for the missing value with subscripts determined by the position of the missing value in the table, e.g. x_{35} is the missing value in

the third block receiving the fifth treatment. Space should be left to insert the estimated missing value.

Table 1. The wet weight in mg (x) of embryonic chick tibiae after cultivation on a chemically defined medium containing five different concentrations of glucose. Treatment numbers are placed in parentheses. The variate analysed is $\log 10x$

Block	Glucose concentration (mg./ml.)					B_i	Completed block totals
	0.5 (1)	1.0 (2)	2.0 (3)	4.0 (4)	8.0 (5)		
I	0.88	1.15	1.33	1.39	1.57	—	6.32
II	1.06	1.22	1.63	1.54	1.37	—	6.82
III	0.97	1.34	$(x_{33} = 1.50)$	1.66	$(x_{35} = 1.53)$	3.97 (B_3)	7.00
IV	1.09	1.21		1.50			6.44
V	1.14	1.37	1.16	1.52	1.48	—	7.05
VI	1.13	1.33	1.58	1.57	1.44	—	7.15
VII	1.00	1.21	1.65	1.45	1.47	—	6.62
VIII	1.12	1.30	$(x_{73} = 1.44)$	$(x_{84} = 1.56)$	1.52	5.18 (B_7)	6.94
					1.61	5.38 (B_8)	
T_j	—	—	8.70 (T_3)	10.63 (T_4)	10.46 (T_5)	48.31 (G)	—
Completed treatment totals	8.39	10.13	11.64	12.19	11.99	—	54.34

Table 2. Calculation of A and q

Subscripts	A				q		
	33	35	73	84	B_i	T_j	G
33	28	—7	—4	1	3.97	8.70	48.31
35	—7	28	1	1	3.97	10.46	48.31
73	—4	1	28	1	5.18	8.70	48.31
84	1	1	1	28	5.38	10.63	48.31
					$r = 8$	$t = 5$	—1

(4) Obtain the B_i , T_j and G from the available observations and write them at the border of the table.

(5) Prepare a table as in Table 2. List the subscripts in a row at the top and a column at the left-hand side of the table. By comparing the subscripts in pairs determine which subscripts are common. This determines the type of association, and the appropriate coefficient is written down at the intersection of the corresponding row and column. At the right of the

table columns for the B_i , T_j and G are prepared. The i subscript in the left-hand column determines the B_i total in a given row, while the j subscript determines the T_j total. Beneath the B_i and T_j columns the values of r and t are written down respectively, and beneath the G column the value -1 is written.

(6) Invert the matrix A . If the exact variances of treatment comparisons are not required A^{-1} need not be obtained explicitly, and it is quicker to solve the equations (2) directly with q determined as in step (7) below (see, for example, Fox (1950)).

(7) Write down the solution as follows:

$$\begin{bmatrix} x_{33} \\ x_{35} \\ x_{73} \\ x_{84} \end{bmatrix} = \begin{bmatrix} 0.039 & 0.010 & 0.005 & -0.002 \\ 0.010 & 0.038 & 0.000 & -0.002 \\ 0.005 & 0.000 & 0.037 & -0.001 \\ -0.002 & -0.002 & -0.001 & 0.036 \end{bmatrix} \begin{bmatrix} 26.95 \\ 35.75 \\ 36.63 \\ 47.88 \end{bmatrix}$$

q is obtained by row by row multiplication of each row of the right-hand side of Table 2 with the bottom row. Thus, in the example, the upper entry of q is given by

$$8 \times 3.97 + 5 \times 8.70 - 48.31 = 26.95.$$

The values of the x_{ij} are then obtained by standard row by column multiplication. In the example $x_{33} = 1.50$; $x_{35} = 1.53$; $x_{73} = 1.44$; $x_{84} = 1.56$.

(8) The values of the x_{ij} are inserted in Table 1, and the block and treatment totals completed. The standard analysis of variance then follows. In this example 4 D.F. are subtracted from the error term.

Special cases. The process of calculating A^{-1} can be simplified if A is a $p \times p$ matrix of the form:

$$\begin{bmatrix} a & b & . & . & . & b \\ b & a & . & . & . & b \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ b & b & . & . & . & a \end{bmatrix}. \quad (5)$$

This is a circulant matrix and hence the sum of the elements of a row is a factor of its determinant. Thus if $\{a + b(p-1)\} = 0$ the matrix is singular. This property will be used later.

The inverse of (5) is given by

$$\frac{1}{(a-b)\{a+b(p-1)\}} \begin{bmatrix} \{a+b(p-2)\} & -b & . & . & . & -b \\ -b & \{a+b(p-2)\} & . & . & . & -b \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ -b & -b & . & . & . & \{a+b(p-2)\} \end{bmatrix}$$

Thus once A is known, A^{-1} can be written down directly.

The special cases where this is useful in the randomized block design are:

- (1) all missing observations are zero associates,
- (2) all missing observations are either block or treatment associates. These matrices also occur in other experimental designs.

III. RANDOMIZED BLOCK DESIGN WITH REPLICATION WITHIN UNITS

Suppose we have r blocks each containing t plots to which t treatments are allotted at random, and that in each plot there are s independent observations. The analysis of variance takes the form shown in Table 3. At the present time there is considerable discussion on the linear models which are assumed in the analysis of this design (Anderson & Bancroft, 1952; Wilk & Kempthorne, 1955). The models determine the expectations of the mean squares in the analysis of Table 3, and these in turn determine the appropriate error terms in the tests of significance. There are three error terms to consider:

- (1) the within-plots sum of squares (E_I) (Case 1),
- (2) the blocks $\times$ treatments sum of squares (E_{II}) (Case 2),
- (3) ($E_I + E_{II}$) (Case 3).

Table 3. Analysis of variance for a $r \times t$ randomized block design with s observations per plot

Source of variation	D.F.	Sum of squares
Between blocks	$(r-1)$	$(B-C)$
Between treatments	$(t-1)$	$(T-C)$
Blocks $\times$ treatments interaction (E_{II})	$(r-1)(t-1)$	$(P-B-T+C)$
Within plots (E_I)	$rt(s-1)$	$(S-P)$
Total	$(rst-1)$	$(S-C)$

Where S = total sum of squares, C = correction for the mean, B = block sum of squares, T = treatment sum of squares, P = plot sum of squares.

We denote the k th observation on the j th treatment in the i th block by

$$X_{ijk} (i = 1, 2, \dots, r; j = 1, 2, \dots, t; k = 1, 2, \dots, s).$$

At the end of the experiment let p_{ij} observations be missing in each plot. Then the total number of missing observations is

$$p = \sum_i \sum_j p_{ij}.$$

The following associates only exist: zero, i , j and ij .

Let P_{ij} = total of the available observations in the ij th plot,

B_i = total of the available observations in the i th block,

T_j = total of the available observations receiving the j th treatment,

G = total of all available observations.

The three sets of equations for estimating the missing values when each of the error terms are minimized have been obtained by an analysis similar to that applied to the randomized block design. The rules which determine the elements of $\mathbf{A}$ and $\mathbf{q}$ for each case are shown in Table 4.

If we write down (2) for these cases, ordering the elements of $\mathbf{A}$ in groups corresponding to the ij , we obtain in general:

$$\begin{bmatrix} A_{11}^{11} & A_{12} & \cdot & \cdot & A_{1,rt} \\ A_{12} & A_{22}^{12} & \cdot & \cdot & A_{2,rt} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ A_{1,rt} & A_{2,rt} & \cdot & \cdot & A_{rt,rt}^{rt} \end{bmatrix} \begin{bmatrix} x^{11} \\ x^{12} \\ \cdot \\ \cdot \\ x^{rt} \end{bmatrix} = \begin{bmatrix} q^{11} \\ q^{12} \\ \cdot \\ \cdot \\ q^{rt} \end{bmatrix}, \quad (6)$$

where $A_{\alpha\beta}^{ij} (\alpha = \beta)$ are submatrices along the main diagonal, one corresponding to each plot.

Each is a square symmetric matrix of order p_{ij} . The superscript denotes the plot, while the subscript denotes the position of the submatrix,

$A_{\alpha\beta} (\alpha \neq \beta)$ are rectangular submatrices,

x^{ij} is a vector of the p_{ij} missing values in plot ij ,

q^{ij} is a vector of p_{ij} elements calculated from the available observations.

Table 4. *Coefficients of associates and typical elements of q for the three cases in the randomized block design with replication within units*

Associate	Coefficient		
	Case 1 (E_I)	Case 2 (E_{II})	Case 3 ($E_I + E_{II}$)
Missing value	$(s-1)$	$(r-1)(t-1)$	$(rst-r-t+1)$
ij	-1	$(r-1)(t-1)$	$(1-r-t)$
i	0	$(1-r)$	$(1-r)$
j	0	$(1-t)$	$(1-t)$
Zero	0	1	1
Typical elements of q			
Case 1	$P_{f\sigma}$ $rB_f + tT_\sigma - rtP_{f\sigma} - G$ $rB_f + tT_\sigma - G$		
Case 2			
Case 3			

If a plot has no missing values the submatrices of the corresponding row and column in (6) vanish. The elements of the submatrices along the main diagonal are determined by the coefficients for the missing value and its ij associates only.

We now consider the three cases discussed above.

Case 1. The matrix of (6) is of the form

$$\begin{bmatrix} A_{11} & \cdot & \cdot & \cdot & \cdot \\ \cdot & A_{12} & \cdot & \mathbf{O} & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \mathbf{O} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & A_{rt,rt} \end{bmatrix} \quad (7)$$

Since the inverse of (7) is

$$\begin{bmatrix} A_{11}^{-1} & \cdot & \cdot & \cdot & \cdot \\ \cdot & A_{12}^{-1} & \cdot & \mathbf{O} & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \mathbf{O} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & A_{rt,rt}^{-1} \end{bmatrix},$$

the estimates of the missing values in plot ij is independent of the missing values in all other plots. Therefore the missing values for the ij th plot are given by the solution of

$$\begin{bmatrix} (s-1) & -1 & . & . & -1 \\ -1 & (s-1) & . & . & -1 \\ . & . & . & . & . \\ . & . & . & . & . \\ -1 & -1 & . & . & (s-1) \end{bmatrix} \begin{bmatrix} x_{ij1} \\ x_{ij2} \\ . \\ . \\ x_{ijp_{ij}} \end{bmatrix} = \begin{bmatrix} P_{ij} \\ P_{ij} \\ . \\ . \\ P_{ij} \end{bmatrix}.$$

$p_{ij} \times p_{ij}$

This has the solution $P_{ij}/(s-p_{ij})$ for all the x_{ijk} . Thus in this case the missing observations are estimated by the average of the available observations in the plot. If all the values in a plot are missing $p_{ij} = s$ and the equations have no straightforward solution.

When $r = 1$ we have the case of a fully randomized design. Thus an estimate of a missing value, if required, is the average of the results available on the particular treatment.

Case 2. The submatrices, in this case, have equal elements throughout. The $A_{\alpha\beta}^{ij}(\alpha = \beta)$ have elements equal to $(r-1)(t-1)$ while the $A_{\alpha\beta}^{ij}(\alpha \neq \beta)$ have elements equal to $(1-r)$, $(1-t)$ or 1. The rank of matrix $\mathbf{A}$ in all practical cases is rt since it consists of rt sets of identical rows of elements. Thus to solve (6) we may allot within each plot $(p_{ij}-1)$ arbitrary values and then solve for the remainder. In practice, therefore, we would insert arbitrary values for all but one missing value in each plot, say $P_{ij}/(s-p_{ij})$ for $s \neq p_{ij}$, and then write out (6) as though one value only is missing in each plot. Alternatively, if we work entirely with plot totals we can use the formula for a randomized block design to estimate a missing plot, ignoring all the available observations on the plot. This situation arises when we minimize the whole-plot error term in a split-plot design.

Case 3. Here the matrix of (6) is non-singular and cannot be simplified. In this case the matrix remains non-singular even if all observations are missing from a plot, thus allowing estimates of these values to be made.

IV. CROSS-OVER DESIGN

Consider an $r \times t$ design, where each of the t columns is a separate replicate of the r treatments, subject to the condition that each treatment occurs an equal number of times (a) in each row. Then $t = ar$.

Denote each observation by $X_{ijk}(i, k = 1, 2, \dots, r; j = 1, 2, \dots, t)$. Let p observations be missing. In this design the following associates only exist: zero, i, j, k, ik .

Let R_i = total of the available observations in the i th row,

C_j = total of available observations on the j th column,

T_k = total of the available observations receiving the k th treatment,

G = total of all available observations.

The following rules determine the elements of $\mathbf{A}$:

- (i) $(r-1)(t-2)$ for the missing observation under consideration,
- (ii) $2(1-r)$ for the ik associates,
- (iii) $(2-r)$ for the i and k associates,
- (iv) $(2-t)$ for the j associates,
- (v) 2 for the zero associates.

A typical element of $\mathbf{q}$ is given by

$$rR_i + tC_j + rT_k - 2G.$$

When $r = 2$ the coefficient for i and k associates is zero. This leads to independence of the equations for estimating the missing values in some instances, thus allowing the formula for a single missing value to be used separately for each plot. A similar situation arises in split-plot designs (see Example 2).

V. LATIN-SQUARE DESIGN

The formulae for a $t \times t$ latin-square design can be obtained from those for the cross-over design by putting $a = 1$. Under this restriction ik associates do not exist. Thus the rules which determine the elements of $\mathbf{A}$ for the latin-square design are:

- (i) $(t-1)(t-2)$ for the missing observation under consideration,
- (ii) $(2-t)$ for the i, j and k associates,
- (iii) 2 for the zero associates.

A typical element of $\mathbf{q}$ is then $t(R_f + C_g + T_h) - 2G$.

VI. HIGHER ORDER SQUARE DESIGNS

Suppose we have a $t \times t$ square of order n , where n denotes the number of classifications. Denote the observation made on each plot by $X_{\alpha_1 \alpha_2 \dots \alpha_n}$ ($\alpha_i = 1, 2, \dots, t$; $i = 1, 2, \dots, n$). A particular observation will be denoted by $X_{\beta_1 \beta_2 \dots \beta_n}$, where the β_i represent fixed values of the α_i . Let p observations be missing.

In all these designs only zero and 1st order associates exist.

Let C_{α_i} = total of the available observations on the α_i th classification,
 G = total of all available observations.

By an extension of the methods already described it can be shown that a typical equation of (2) is

$$\begin{aligned}
 & (t-1)(t+1-n)x_{\beta_1 \beta_2 \dots \beta_n} + (n-1-t) \sum_{\substack{\alpha_2 \neq \beta_2 \\ \alpha_3 \neq \beta_3}} \sum_{\substack{\alpha_3 \neq \beta_3 \\ \alpha_4 \neq \beta_4}} \dots \sum_{\alpha_n \neq \beta_n} x_{\beta_1 \alpha_2 \dots \alpha_n} \\
 & + (n-1-t) \sum_{\substack{\alpha_1 \neq \beta_1 \\ \alpha_3 \neq \beta_3}} \sum_{\substack{\alpha_3 \neq \beta_3 \\ \alpha_4 \neq \beta_4}} \dots \sum_{\alpha_n \neq \beta_n} x_{\alpha_1 \beta_2 \dots \alpha_n} \\
 & + \dots \dots \dots \\
 & + (n-1-t) \sum_{\substack{\alpha_1 \neq \beta_1 \\ \alpha_2 \neq \beta_2}} \sum_{\substack{\alpha_2 \neq \beta_2 \\ \alpha_{n-1} \neq \beta_{n-1}}} \dots \sum_{\alpha_n \neq \beta_n} x_{\alpha_1 \alpha_2 \dots \beta_n} \\
 & + (n-1) \sum_{\substack{\alpha_1 \neq \beta_1 \\ \alpha_2 \neq \beta_2}} \sum_{\alpha_3 \neq \beta_3} \dots \sum_{\alpha_n \neq \beta_n} x_{\alpha_1 \alpha_2 \dots \alpha_n} \\
 & = tC_{\beta_1} + tC_{\beta_2} + \dots + tC_{\beta_n} - (n-1)G.
 \end{aligned} \tag{8}$$

From (8) we may determine the rules shown in Table 5 for writing down the coefficients of matrix $\mathbf{A}$ for various values of n . $n = 2$ determines a randomized block design where the number of treatments equals the number of blocks. $n = 3$ determines the coefficients of a $t \times t$ latin square, thus giving an alternative proof of the formula given earlier.

If in the higher square designs $(t-1)$ 1st order associates only are missing then matrix $\mathbf{A}$ is of the form (5). The sum of the elements in a row is

$$(t-1)(t-n+1) + (t-1)(n-1-t) = 0,$$

and therefore the matrix $\mathbf{A}$ is singular and equation (2) has no straightforward solution. This situation arises whenever all the missing values are in one classification, e.g. a complete row, column or treatment in a latin square. A full description of methods for dealing with missing values in latin-square designs where $\mathbf{A}$ is singular is given by De Lury (1946).

Table 5. Coefficients of associates for square designs of different order

No. of constraints	2	3	4	5	n
Associate	Randomized block $s = r = t$	Latin square	Graeco-latin square	Hyper-graeco-latin square	—
Missing observation	$(t-1)(t-1)$	$(t-1)(t-2)$	$(t-1)(t-3)$	$(t-1)(t-4)$	$(t-1)(t-n+1)$
1st order	$(1-t)$	$(2-t)$	$(3-t)$	$(4-t)$	$(n-1-t)$
Zero	1	2	3	4	$(n-1)$

VII. SPLIT-PLOT DESIGNS

(1) Whole-plot treatments arranged in randomized blocks

Consider a design in which r blocks containing t plots to which t whole-plot treatments are allotted at random. Further let s split-plot treatments be allotted at random within each whole plot. The analysis of variance of this design is given in Table 6. There are two error terms to be considered: (1) E_I corresponding to the whole-plots (Case 1), and (2) E_{II} corresponding to the split-plots (Case 2).

Table 6. Analysis of variance for a split-plot design with whole-plot treatments arranged in randomized blocks

Source of variation	D.F.	Sum of squares
Between blocks	$(r-1)$	$(B-C)$
Between whole-plot treatments (A)	$(t-1)$	$(S-C)$
Error (E_I)	$(r-1)(t-1)$	$(W-B-S+C)$
Total	$(rt-1)$	$(W-C)$
Between split-plot treatments (B)	$(s-1)$	$(T-C)$
AB interaction	$(s-1)(t-1)$	$(I-S-T+C)$
Error (E_{II})	$t(s-1)(r-1)$	$(Z-W-I+S)$
Total	$(rst-1)$	$(Z-C)$

Where Z = total sum of squares of split-plots, W = total sum of squares of whole plots, C = correction for the mean, B = block sum of squares, S = whole-plot treatment sum of squares, T = split-plot treatment sum of squares, I = treatment combination sum of squares.

We denote the observation made on the k th split-plot treatment in the ij th plot by X_{ijk} ($i = 1, 2, \dots, r; j = 1, 2, \dots, t; k = 1, 2, \dots, s$). At the end of the experiment let p split-plots be missing.

The following associates exist: zero, i, j, k, ij, ik, jk .

Case 1. Table 6 shows that the expression for the error term is the same as that in Case 2 of the generalized randomized block design. The solution in the two cases is identical. We ignore the available observations and estimate new whole-plot values using the formula for a randomized block design.

Case 2. Let W_{ij} = total of the available observations in the ij th plot,
 I_{jk} = total of the available observations receiving the j th whole-plot treatment and the k th split-plot treatment,
 S_j = total of the available observations receiving the j th whole-plot treatment.

The rules which determine the elements of $\mathbf{A}$ are:

- (i) $(r-1)(s-1)$ for the missing observation under consideration,
- (ii) $(1-s)$ for the jk associates,
- (iii) $(1-r)$ for the ij associates,
- (iv) 1 for the j associates,
- (v) 0 for the zero i , k and ik associates.

A typical element of $\mathbf{q}$ is given by:

$$rW_{fg} + sI_{gh} - S_g.$$

(2) Whole-plot treatments in a $t \times t$ latin square

Consider a $t \times t$ latin square to which whole-plot treatments are allotted at random, and further let s split-plot treatments be allotted within each whole plot. The analysis of variance is shown in Table 7. As before there are two error terms to be considered.

Table 7. Analysis of variance for a split-plot design with whole-plot treatments arranged in a latin square

Source of variation	D.F.	Sum of squares
Between rows	$(t-1)$	$(R-C)$
Between columns	$(t-1)$	$(K-C)$
Between whole-plot treatments (A)	$(t-1)$	$(S-C)$
Error (E_I)	$(t-1)(t-2)$	$(W-R-K-S+2C)$
Total	(t^2-1)	$(W-C)$
Between split-plot treatments (B)	$(s-1)$	$(T-C)$
AB interaction	$(t-1)(s-1)$	$(I-S-T+C)$
Error (E_{II})	$t(t-1)(s-1)$	$(Z-W-I+S)$
Total	(st^2-1)	$(Z-C)$

Where Z = total sum of squares of split-plots, W = total sum of squares of whole plots, C = correction for the mean, R = row sum of squares, K = column sum of squares, S = whole-plot treatment sum of squares, T = split-plot treatment sum of squares, I = treatment combination sum of squares.

We denote an observation on each split-plot by $X_{ijkl}(i, j, k = 1, 2, \dots, t; l = 1, 2, \dots, s)$, where i denotes the row, j the column, k the whole-plot treatment associated with the plot ij , and l the split-plot treatment. Let p split-plots be missing.

The following associates only exist due to the restrictions on a latin-square design: zero, $i, j, k, l, il, jl, kl, ijk$.

Case 1. If we wish to minimize the whole-plot error we ignore the available observations and estimate new whole-plot values using the formula for a latin-square design.

Case 2. Let W_{ij} = total of the available observations on the ij th plot,

I_{kl} = total of the available observations receiving the k th whole-plot treatment and the l th split-plot treatment,

S_k = total of the available observations receiving the k th whole-plot treatment.

The rules which determine the elements of $\mathbf{A}$ are:

- (i) $(t-1)(s-1)$ for the missing observation under consideration,
- (ii) $(1-t)$ for the ijk associates,
- (iii) $(1-s)$ for the kl associates,
- (iv) 1 for the k associates,
- (v) 0 for the zero, i, j, l, il and jl associates.

A typical element of $\mathbf{q}$ is given by $tW_{ef} + sI_{gh} - S_g$.

The fact that the coefficient of some associates is zero in these split-plot designs allows the problem of solving the equations to be simplified in many cases. If the matrix $\mathbf{A}$ is written down ordering the elements in groups according to the j subscript for the randomized block design and the k subscript for the latin-square design a matrix of the form (7) is obtained. This consists of submatrices along the main diagonal, one corresponding to each j or k , and zeros elsewhere. As these submatrices may be separately inverted to give the required solution considerable reduction in the labour of computation results. Further, in many simple cases the missing observations may be fitted quite independently using the standard formula for one missing observation. The independence of missing values within whole plots in these designs has been pointed out previously by Pearce (1953). If an entire whole plot is missing the $t \times t$ submatrix will have $(r-1)(t-1)$ along the main diagonal and $(1-r)$ elsewhere. The sum of a row is zero, and thus the submatrix is singular and there is no straightforward solution to the equations.

Example 2. Table 8 shows the results of a split-plot experiment where the whole-plot treatments are arranged in randomized blocks ($r = 4; t = 5; s = 2$). Missing values have been artificially chosen to provide the example. Table 9 is prepared from the data of Table 8 in exactly the same way as described in Example 1, except that the subscripts have been ordered using the middle (j) subscript. From the matrix $\mathbf{A}$ it is seen that the estimation of x_{121} and x_{222} is independent of the estimation of x_{132} and x_{232} . Thus instead of inverting of a 4×4 matrix we invert two 2×2 matrices, considerably simplifying the problem.

The estimates are given by

$$\begin{aligned} \begin{bmatrix} x_{121} \\ x_{222} \end{bmatrix} &= 1/8 \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 4.74 \\ 4.74 \end{bmatrix} \\ \text{and} \quad \begin{bmatrix} x_{132} \\ x_{232} \end{bmatrix} &= 1/8 \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 2.56 \\ 3.76 \end{bmatrix}. \end{aligned}$$

Thus $x_{121} = 1.19; x_{222} = 1.19; x_{132} = 1.43; x_{232} = 1.73$.

Table 8. *The wet weight in mg (x) of embryonic chick tibiae after cultivation on a chemically defined medium containing different concentrations of glucose or mannose. Treatment numbers are placed in parentheses. The variate analysed is $\log 10x$*

Block	Split-plot treatment	Whole-plot treatment Hexose concentration (mg./ml.)				
		0.5 (1)	1.0 (2)	2.0 (3)	4.0 (4)	8.0 (5)
I	Glucose (1)	0.88	$(x_{121} = 1.19)$ 1.15	1.33	1.39	1.57
	Mannose (2)	0.78		$(x_{132} = 1.43)$ 1.42	1.42	1.51
II	Glucose (1)	1.06	1.22	1.63	1.54	1.37
	Mannose (2)	1.09	$(x_{222} = 1.19)$ 1.19	$(x_{232} = 1.73)$ 1.47	1.47	1.44
III	Glucose (1)	0.97	1.34	1.44	1.66	1.59
	Mannose (2)	0.90	1.16	1.44	1.48	1.59
IV	Glucose (1)	1.09	1.21	1.16	1.50	1.48
	Mannose (2)	1.07	1.32	1.36	1.41	1.41

Table 9. *Calculation of A and q*

Subscripts	A				q		
	121	222	132	232	W_{ij}	I_{jk}	S_j
121	3	1	0	0	1.15	3.77	7.40
222	1	3	0	0	1.22	3.63	7.40
132	0	0	3	-1	1.33	2.80	8.36
232	0	0	-1	3	1.63	2.80	8.36
					$r = 4$	$s = 2$	-1

VIII. MIXED-UP OBSERVATIONS

Sometimes the identity of observations may be lost so that we only know their total. These are known as mixed-up values. Bose & Mahalanobis (1938) and Nair (1940) have discussed the analysis of randomized block and latin-square designs where some observations are mixed up, and they have shown that estimates may be obtained which minimize the error mean square.

In the general case both missing and mixed-up values may occur. We may relate this to the previous sections by supposing that some of the p missing observations are not missing entirely, but that certain of them have been mixed up. Then the p missing values may be connected by c relations of the form:

$$\sum_{(i)} b_{ij} x_i = u_j \quad (i = 1, 2, \dots, p; j = 1, 2, \dots, c) \quad (9)$$

The coefficients of (9) are either 0 or 1. For example, if $p = 5$, and the missing values are x_2, x_7, x_9, x_{13} and x_{19} , with x_7 and x_9 and also x_2 and x_{19} mixed up and x_{13} entirely missing, we have the relations:

$$0.x_2 + 1.x_7 + 1.x_9 + 0.x_{13} + 0.x_{19} = u_1$$

and

$$1.x_2 + 0.x_7 + 0.x_9 + 0.x_{13} + 1.x_{19} = u_2.$$

To obtain estimates for the missing and mixed-up values we minimize the error mean square subject to the conditional equations (9). Although the coefficients of (9) are restricted to 0 and 1 to deal with the case of mixed-up values, this restriction is not essential in the following discussion. The problem will be treated generally in matrix notation.

Equation (8) may be written:

$$\mathbf{B}\mathbf{x} = \mathbf{u}, \quad (10)$$

where $\mathbf{B}$ is a $p \times c$ matrix of coefficients; each column of this matrix will contain one unit element and $(c - 1)$ zero elements, $\mathbf{x}$ is a column vector of the p values to be estimated, $\mathbf{u}$ is a column vector of c mixed-up totals.

If we introduce a column vector $(2\lambda/N)$ of c undetermined multipliers the Lagrange equations are given by

$$\frac{\partial}{\partial \mathbf{x}} \{F(x_1, x_2, \dots, x_p)\} + 2\lambda'(\mathbf{B}\mathbf{x} - \mathbf{u}) = \mathbf{O}.$$

Using the notation of equation (1) we find that

$$2/N(\mathbf{x}'\mathbf{A} - \mathbf{q}') + 2/N\lambda'\mathbf{B} = \mathbf{O},$$

and after transposition and rearrangement

$$\mathbf{A}\mathbf{x} + \mathbf{B}'\lambda = \mathbf{q}. \quad (11)$$

The values of the p missing values which minimize the error mean square are given by the simultaneous solution of (10) and (11). These two matrix equations may be combined to give

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}' \\ \mathbf{B} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{q} \\ \mathbf{u} \end{bmatrix},$$

the solution of which is given by:

$$\begin{bmatrix} \mathbf{x} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{B}' \\ \mathbf{B} & \mathbf{O} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{q} \\ \mathbf{u} \end{bmatrix}.$$

Example 3. Suppose that in Example 2, x_{121} and x_{132} have been confused and that the combined weights after transformation to logs is 2.50. Similarly, suppose that x_{222} and x_{232} are also confused, and that the combined weight is 2.83.

The matrix of coefficients is found to be

$$\begin{array}{cc} \mathbf{A} & \begin{bmatrix} 3 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 0 & 0 & 3 & -1 \\ 0 & 0 & -1 & 3 \end{bmatrix} \\ \mathbf{B} & \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \mathbf{B}' \\ \mathbf{O} \end{array} \begin{array}{c} \mathbf{q} \\ \mathbf{u} \end{array}$$

Then the missing observations are given by

$$\begin{bmatrix} x_{121} \\ x_{222} \\ x_{132} \\ x_{232} \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 1 & 0 & -1 & 0 & 3 & -1 \\ 0 & 1 & 0 & -1 & -1 & 3 \\ -1 & 0 & 1 & 0 & 3 & 1 \\ 0 & -1 & 0 & 1 & 1 & 3 \\ 3 & -1 & 3 & 1 & -8 & 0 \\ -1 & 3 & 1 & 3 & 0 & -8 \end{bmatrix} \begin{bmatrix} 4.74 \\ 4.74 \\ 2.56 \\ 3.76 \\ 2.50 \\ 2.83 \end{bmatrix}.$$

Thus $x_{121} = 1.14$; $x_{222} = 1.16$; $x_{132} = 1.36$; $x_{232} = 1.67$.

I wish to thank Dr D. R. Cox for helpful comments on the draft of this paper.*

REFERENCES

- ALLAN, F. E. & WISHART, J. (1930). *J. Agric. Sci.* **20**, 399.
 ANDERSON, R. L. (1946). *Biometrics*, **2**, 41.
 ANDERSON, R. L. & BANCROFT, T. A. (1952). *Statistical Theory in Research*. London: McGraw-Hill.
 BARTLETT, M. S. (1937). *J.R. Statist. Soc. Suppl.* **4**, 137.
 BATEN, W. D. (1939). *Technical Bulletin* 165, Michigan State College, Agricultural Experiment Station.
 BATEN, W. D. (1952). *Biometrics*, **8**, 42.
 BOSE, S. S. & MAHALANOBIS, P. C. (1938). *Sankhyā*, **4**, 103.
 COONS, I. (1957). *Biometrics*, **13**, 387.
 DE LURY, D. B. (1946). *J. Amer. Statist. Ass.* **41**, 370.
 FEDERER, W. T. (1951). *Research Bulletin* 380, Iowa State College, Agricultural Experiment Station.
 FEDERER, W. T. (1955). *Experimental Design*. New York: Macmillan and Co.
 FOX, L. (1950). *J.R. Statist. Soc. B*, **12**, 120.
 FOX, L. & HAYES, J. G. (1951). *J.R. Statist. Soc. B*, **13**, 83.
 HEALY, M. & WESTMACOTT, M. (1956). *Appl. Statist.* **5**, 203.
 KHARGONKAR, S. A. (1948). *J. Indian Soc. Agric. Statist.* **1**, 147.
 NAIR, K. R. (1940). *Sankhyā*, **4**, 581.
 PEARCE, S. C. (1953). *Field Experimentation with Fruit Trees and other Perennial Plants*. Commonwealth Agricultural Bureau.
 THOMPSON, H. R. (1956). *Biometrics*, **12**, 241.
 WILK, M. B. & KEMPTHORNE, O. (1955). *J. Amer. Statist. Ass.* **50**, 1144.
 WILKINSON, G. N. (1957). *Biometrics*, **13**, 363.
 WILKINSON, G. N. (1958a). *Biometrics*, **14**, 257.
 WILKINSON, G. N. (1958b). *Biometrics*, **14**, 360.
 YATES, F. (1933). *Emp. J. Exp. Agric.* **1**, 129.

* Since this paper was submitted for publication two further papers have appeared by Wilkinson (1958a, b) in which the problem of estimating missing values is discussed more generally. In these papers Wilkinson also deals with missing values in balanced and partially balanced non-orthogonal designs, and discusses the analysis of those cases where $\mathbf{A}$ is singular.

NOTE ON A THREE-DECISION TEST FOR COMPARING TWO BINOMIAL POPULATIONS

By J. H. DARWIN

*Department of Scientific and Industrial Research, Wellington,
New Zealand*

1. INTRODUCTION AND SUMMARY

Suppose that in each of two processes there is a probability of an event happening and it is desired to compare these probabilities. Pairs of trials of these processes may be made in sequence and a test made on the results. There may be four possible outcomes of each pair of trials—the event may happen in both processes, in neither, or in only one. If the result of the test is to be that just one of the processes will be used in the future the first two outcomes are irrelevant. If the result of the test is to be that just one of the processes will be used in the future if it is markedly superior to the other, but that either may be used if they do not appear to differ importantly then all outcomes are relevant.

In this paper we produce a test for this latter case. Another test of this kind has recently been described by Armitage (1957). He does not employ pairs of trials having the same outcome and consequently has a different form of test. Both his and our tests are truncated in the sense that a maximum number of pairs of trials is needed for each.

When the probability of the same result for each process is well known, our test will require a smaller maximum number of pairs of trials for the same power, and the same probability of rejecting when true the null hypothesis that the processes are equally effective. A numerical example of this is given.

When this probability of the same result for each process is not known Armitage's test is preferable since it does not depend on a knowledge of this probability. When, however, the tests are being used to compare a control process with a new process our test which, in its construction assumes some value of the probability of the same result, will be affected only favourably should the control value of this probability depart from the assumed value.

The strongest connexion between our test and Armitage's is that the boundaries which determine when our test stops are a special case of the boundaries required by his test. We use exact matrix theory to find the probabilities of the boundaries being crossed, and he uses approximate normal theory. For some comparative examples we calculated, the probability given by this approximate theory differed from those given by the exact theory by at most 0.0015.

We also give an approximate set of power contours which enable tests of our type to be constructed for any value of the power and of the probability of rejecting the null hypothesis when it is true.

2. DERIVATION OF THE TEST

Suppose the probabilities of the events happening are p_1 and p_2 for the two processes.

Then in a pair of trials the probability of the event happening in the first process but not in the second is $p_1(1-p_2) = \pi_1$, say.

The probability of the same result in the two trials is $p_1 p_2 + (1-p_1)(1-p_2) = \pi_2$, say.

The probability of the event happening in the second process but not in the first is $p_2(1-p_1) = \pi_3$, say.

We consider a variable γ taking values 1, 0 and -1 for these three cases, respectively. Suppose m pairs of trials have been made in sequence and $\gamma(m)$ is the sum of the m values of γ . We could imagine $\gamma(m)$ drawn on a graph as a function of m . Suppose $\gamma(0) = \frac{1}{2}(n+1)$ where n is an odd integer, and suppose the probability that $\gamma(m)$ has the value Y is $P_{Y,m}$. Then

$$P_{Y,m+1} = \pi_1 P_{Y-1,m} + \pi_2 P_{Y,m} + \pi_3 P_{Y+1,m}. \quad (1)$$

If $\gamma(m)$ crosses $Y = 0$ we shall suppose the first process is accepted as that to use in the future. If $\gamma(m)$ crosses $Y = n+1$ we shall suppose the second process is accepted. If $\gamma(m)$ has crossed neither of these boundaries when M pairs of trials have been made the processes will be regarded as equally acceptable for future use and the sequence stopped. We now wish to find values of M and n for which there is a high probability of a markedly superior process being accepted, and a high probability of either process being permitted in the future if there is the same probability of the event happening in each process. To find these values we solve the system (1) with $P_{1-a,m} = P_{n+b,m} = 0$ for all integers a and b greater than 0; i.e. for $\gamma(m)$ confined between 1 and n . The $P_{Y,m}$ satisfy the matrix equation

$$\mathbf{P}_{m+1} = \begin{pmatrix} \pi_2 & \pi_3 & \cdot & \cdot & \cdot & \cdot \\ \pi_1 & \pi_2 & \pi_3 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \pi_1 & \pi_2 & \pi_3 \\ \cdot & \cdot & \cdot & \cdot & \pi_1 & \pi_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \mathbf{P}_m$$

$$= D_n \mathbf{P}_m, \quad \text{say,}$$

where $\mathbf{P}_m$ is the vector of the $P_{Y,m}$.

Suppose $|D_n - \lambda I|$ is called $D_n(\lambda)$. Then $D_u(\lambda)$ satisfies the difference equation

$$D_u(\lambda) = (\pi_2 - \lambda) D_{u-1}(\lambda) - \pi_1 \pi_3 D_{u-2}(\lambda)$$

with solution

$$D_u(\lambda) = \frac{(\frac{1}{2}[\pi_2 - \lambda + \{(\pi_2 - \lambda)^2 - 4\pi_1 \pi_3\}^{\frac{1}{2}}])^{u+1} - (\frac{1}{2}[\pi_2 - \lambda - \{(\pi_2 - \lambda)^2 - 4\pi_1 \pi_3\}^{\frac{1}{2}}])^{u+1}}{\{(\pi_2 - \lambda)^2 - 4\pi_1 \pi_3\}^{\frac{1}{2}}}.$$

It easily follows that the latent roots of D_n are

$$\lambda_r = \pi_2 + 2\sqrt{(\pi_1 \pi_3)} \cos \frac{r\pi}{n+1} \quad (r = 1, 2, \dots, n). \quad (3)$$

It then follows that

$$D_u(\lambda_r) = (\pi_1 \pi_3)^{\frac{1}{2}u} \sin [(u+1)r\pi/(n+1)].$$

The elements of the latent column vector C_r corresponding to λ_r satisfy the equation

$$\pi_1 v_j + (\pi_2 - \lambda) v_{j+1} + \pi_3 v_{j+2} = 0.$$

It then follows that the j th element of C_r is proportional to $(-)^j D_{j-1}(\lambda_r)/\pi_3^j$; and, similarly, the j th element of the latent row vector R_r corresponding to λ_r is proportional to $(-)^j D_{j-1}(\lambda_r)/\pi_1^j$.

A spectral resolution of the matrix D_n is then possible (see e.g. Bartlett 1955, p. 26) and we may write

$$D_n^m = \sum_{r=1}^n \lambda_r^m \frac{C_r R_r'}{C_r' R_r}.$$

The probability of $\gamma(m)$ crossing $Y = 0$ up to the time when M pairs of trials have been made, if $\gamma(m)$ has previously been between 1 and n , is

$$\sum_{m=0}^{M-1} P_{1m} \pi_3 = \pi_3 \sum_m \sum_{r=1}^n \lambda_r^m C_{r,1} R_{r, \frac{1}{2}(n+1)} / (C'_r R_r)$$
$$= \frac{2}{n+1} \pi_3 \left(\frac{\pi_3}{\pi_1} \right)^{\frac{1}{4}(n-1)} \sum_{r=1}^n \frac{1 - \lambda_r^M}{1 - \lambda_r} \sin \frac{r\pi}{n+1} \sin \frac{r\pi}{2}.$$

(5)

Similarly, the probability of $\gamma(m)$ crossing the boundary $Y = n + 1$ up to the time when M pairs of trials have been made, if $\gamma(m)$ has previously been between 1 and n , is

$$\frac{2}{n+1} \pi_1 \left(\frac{\pi_1}{\pi_3} \right)^{\frac{1}{4}(n-1)} \sum_{r=1}^n \frac{1 - \lambda_r^M}{1 - \lambda_r} \sin \frac{r\pi}{n+1} \sin \frac{r\pi}{2}.$$

(6)

The probability of $\gamma(m)$ crossing the line $m = M$ between $Y = 1$ and $Y = n$ without having previously crossed either $Y = 0$ or $Y = n + 1$ is

$$\frac{2}{n+1} \left(\pi_3 \left(\frac{\pi_3}{\pi_1} \right)^{\frac{1}{4}(n-1)} + \pi_1 \left(\frac{\pi_1}{\pi_3} \right)^{\frac{1}{4}(n-1)} \right) \sum_{r=1}^n \frac{\lambda_r^M}{1 - \lambda_r} \sin \frac{r\pi}{n+1} \sin \frac{r\pi}{2}.$$

(7)

3. CALCULATIONS AND DISCUSSION

Values of M and n given in Table 1 have been found to satisfy these requirements:

- (i) For some π_2 and for a value of π_1/π_3 greater than 1, the probability (6) was made equal to 0.95 or 0.80.

(8)
- (ii) For the same π_2 and for $\pi_1 = \pi_3$ the probability (7) was made equal to 0.95.

(8')

That is, the probability of accepting, when true, the null hypothesis of no difference between the processes, is 0.95, and the power of the test when π_1 and π_3 differ is either 0.95 or

Table 1. Distance between acceptance limits (n) and number of trials (M) which will ensure the power indicated, at the 5 % significance level

	Power	$\pi_1/\pi_3 (= S) = 1.2$		1.4	1.6	1.8	2.0	2.5	3.0
$\pi_2 = 0.25$	0.95	n	181	99	71	57	49	37	31
		M	2200	656	340	2200	162	95	68
	0.80	n	141	77	55	45 (43)	37	29	25
		M	1336	399	206	136 (131)	96	59	43
$\pi_2 = 0.50$	0.95	n	181	99	71	57	49	37	31
		M	3301	986	511	331	244	144	103
	0.80	n	141	77	55	45	37	29	25
		M	2006	600	311	205	145	89	65
$\pi_2 = 0.75$	0.95	n	181	99	71	57	49	37	31
		M	6606	1972	1025	666	491	290	209
	0.80	n	141	77	55	45	37	29	25
		M	4018	1201	623	412	292	179	132

0.80. The conditions (8) determine unique values of M and n . The range of π_1/π_3 considered was 1 to 3.

That value of n was chosen for which the difference between the values of M for which (i) and (ii) of (8) were satisfied was a minimum. Then for the sake of definiteness that value of M was taken for which the power was closest to 0.95 or 0.80. The two values of n given at one place under $\pi_1/\pi_3 = 1.8$ indicate that $n = 43$, $M = 131$ is the better choice, but $n = 45$, $M = 136$ has been put in for comparison with the lower figures in this column.

The calculations were not as laborious as might be imagined, since in all cases at most six and usually three terms of the series in (7) had to be taken, and always for $\pi_1 > \pi_3$, (5) was negligible.

In (6)

$$\frac{2}{n+1} \pi_1 \left(\frac{\pi_1}{\pi_3} \right)^{\frac{1}{2}(n-1)} \sum_r \frac{\sin(r\pi/(n+1)) \sin \frac{1}{2} r\pi}{1 - \lambda_r} = \frac{\pi_1^{\frac{1}{2}(n+1)}}{\pi_1^{\frac{1}{2}(n+1)} + \pi_3^{\frac{1}{2}(n+1)}}, \quad (9)$$

since this is the known probability of $\gamma(m)$ crossing $Y = n+1$ when M is infinite. We used this as a partial check on the sufficiency of the numbers of figures we were keeping.

There are good approximations available to the probabilities (6) and (7). First, there is the normal approximation. In this $\gamma(M)$ for large M is taken as normally distributed with mean $M(\pi_1 - \pi_3)$ and variance $M(1 - \pi_2 - (\pi_1 - \pi_3)^2)$. The probability of $\gamma(m)$ crossing the boundary $Y = n+1$ before M pairs of trials have been made is found (after consideration of the paths that reach $m = M$ with Y less than $n+1$ but which have previously crossed $Y = n+1$) to be

$$1 - F \left[\frac{\frac{1}{2}(n+1) - (\pi_1 - \pi_3) M}{\{M(1 - \pi_2 - (\pi_1 - \pi_3)^2)\}^{\frac{1}{2}}} \right] + \exp \left[\frac{(n+1)(\pi_1 - \pi_3)}{1 - \pi_2 - (\pi_1 - \pi_3)^2} \right] F \left[\frac{-\frac{1}{2}(n+1) - (\pi_1 - \pi_3) M}{\{M(1 - \pi_2 - (\pi_1 - \pi_3)^2)\}^{\frac{1}{2}}} \right], \quad (10)$$

where

$$F(x) = \int_{-\infty}^x e^{-\frac{1}{2}x^2} dx / \sqrt{(2\pi)}.$$

This result is quoted by Armitage (1957) from work by Bartlett (1946). We compared the powers it gave for the values of n and M given by the first part of the line for $\pi_2 = 0.75$ in Table 1 with the powers given by the exact quantity (6) and in none of the seven cases was the difference as much as 0.0015.

Another approach to an approximation is given by considering (7) directly. Because of (9) we may write (7) in the form

$$\frac{1}{n+1} \sum_r \frac{(1 - \pi_2) \lambda_r^M \sin \frac{r\pi}{n+1} \sin \frac{r\pi}{2}}{1 - \lambda_r} = \frac{1}{n+1} \sum_r \frac{(1 - \pi_2) \sin \frac{r\pi}{n+1} \sin \frac{r\pi}{2}}{1 - \lambda_r}. \quad (12)$$

Now suppose for convenience we denote π_1/π_3 by S . Then,

$$1 - \lambda_r = (1 - \pi_2) \left(1 - \frac{2\sqrt{S}}{1+S} \cos \frac{r\pi}{n+1} \right). \quad (13)$$

For small r and large n we may write this as

$$\begin{aligned} & \frac{1 - \pi_2}{(n+1)^2} \left(\left(1 - \frac{2\sqrt{S}}{1+S} \right) (n+1)^2 + \frac{2\sqrt{S} r^2 \pi^2}{1+S} \right) \\ &= \frac{1 - \pi_2}{(n+1)^2} \left(\left(1 - \frac{2\sqrt{S}}{1+S} \right) (n+1)^2 + \frac{r^2 \pi^2}{2} \right) \left(1 - \frac{(1 - 2\sqrt{S}/(1+S))^{\frac{1}{2}} r^2 \pi^2}{(1 - 2\sqrt{S}/(1+S)) (n+1)^2 + \frac{1}{2} r^2 \pi^2} \right). \end{aligned} \quad (14)$$

For the range of S considered, 3 to 1.2, $1 - 2\sqrt{S}/(1+S)$ goes from 0.8660 to 0.9959. For the values of n and r encountered this meant that the last factor of (14) departs from one by only a small amount. We only had to consider values of r less than 7 so that r had a range virtually independent of n . This means we may write the denominator of (12) approximately as

$$\sum_{r=1}^6 \frac{r\pi \sin(\frac{1}{2}r\pi)}{(n+1)^2 (1 - 2\sqrt{S}/(1+S) + \frac{1}{2}r^2\pi^2)}. \quad (15)$$

This exhibits the denominator as a function of $(n+1)^2 (1 - 2\sqrt{S}/(1+S))$ only. We shall use this approximation below. Again, $M \log \lambda_r$ for low r is approximately

$$M \log (\pi_2 + 2\sqrt{(\pi_1 \pi_3)}) - \frac{r^2 \pi^2 2M\sqrt{(\pi_1 \pi_3)}}{2(n+2)^2 (\pi_2 + 2\sqrt{(\pi_1 \pi_3)})} \quad (16)$$

$$\simeq -M(1-\pi_2)(1 - 2\sqrt{S}/(1+S)) - \frac{1}{2}M(1-\pi_2)^2 (1 - 2\sqrt{S}/(1+S))^2 - \frac{r^2 \pi^2 M(1-\pi_2)}{2(n+1)^2} \frac{2\sqrt{S}}{1+S} (1 + (1-\pi_2)(1 - 2\sqrt{S}/(1+S))). \quad (17)$$

The two leading terms of (17) each have M and π_2 occurring only in the combination $M(1-\pi_2)$. The remaining terms are proportional to $1 - 2\sqrt{S}/(1+S)$. In particular if $S = 1$, $1 - 2\sqrt{S}/(1+S) = 0$, and the approximation given by the leading terms is especially good. We can in fact write approximately for (7)

$$\frac{4}{\pi} \left(e^{-f} - \frac{e^{-9f}}{3} + \frac{e^{-25f}}{5} - \dots \right), \quad (18)$$

where $f = \pi^2 M(1-\pi_2)/(n+1)^2$. The value of f making (18) equal to 0.95 is 0.24557.

Suppose now we have found M and n for the test to have a given power for a high value of π_2 , say 0.75, and for S not equal to one. Then $M \log \lambda_r(0.75)$ for $\pi_2 = 0.75$ will be larger than $M_1 \log \lambda_r(0.25)$ for $\pi_2 = 0.25$ and $M_1 = \frac{1}{3}M$ because both the second-order approximations in (17) have a negative sign. However $(M_1 - 1) \log \lambda_r(0.25) - M_1 \log \lambda_r(0.25)$ is approximately $0.75(1 - 2\sqrt{S}/(1+S)) + r^2 \pi^2 0.75 \sqrt{S}/\{(1+S)(n+1)^2\}$ and this in our examples is of the same order as the second terms in (17). In fact a drop of only one or two from M_1 was needed to balance the negative terms. This explains why in Table 1, for fixed S and n we are able to achieve the same power when $M(1-\pi_2)$ is kept approximately constant.

We may write the term

$$\frac{r^2 \pi^2 M(1-\pi_2)}{2(n+1)^2} \frac{2\sqrt{S}}{1+S}$$

of (17) as

$$\frac{r^2 \pi^2 M(1-\pi_2)}{2(n+1)^2} \left(1 - \left(1 - \frac{2\sqrt{S}}{1+S} \right) \right). \quad (19)$$

The term in $1 - 2\sqrt{S}/(1+S)$ of (19) is of the same order as the second-order term in (17). It now appears from (15), (17) and (19) that the power (the complement of (7)) is approximately a function of the two quantities

$$\frac{M(1-\pi_2)}{(n+1)^2} \quad \text{and} \quad (n+1)^2 (1 - 2\sqrt{S}/(1+S)). \quad (20)$$

The second-order term in (17) and those neglected in (15) offset each other to some extent. If they are neglected we may, following (20), solve approximately for the M and n required by (8), the equations

$$\frac{M(1-\pi_2)}{(n+1)^2} = \text{constant},$$

$$(n+1)^2 (1 - 2\sqrt{S}/(1+S)) = \text{constant}.$$

For the values of n and S from Table 1 for a power of 0.95 and for increasing n , $(n+1)^2(1-2\sqrt{S/(1+S)})$ takes values 137.19, 139.33, 142.98, 140.23, 139.93, 139.87 and 137.17. For a power of 0.80 the values are 90.57, 86.86, 82.58, 88.21, 84.65, 85.10 and 83.50.

For $\pi_2 = 0.75$, a power of 0.95 and for increasing n , $M(1-\pi_2)/(n+1)^2$ takes values 0.0510, 0.0502, 0.0491, 0.0495, 0.0494, 0.0493 and 0.0499. For a power of 0.80 the values are 0.0488, 0.0497, 0.0506, 0.0487, 0.0497, 0.0494 and 0.0498.

The variations in these quantities are caused partly by the influence of the term neglected in arriving at (20), but mainly by the fact that M and n have to be integers and n has to be odd. This means that the powers for the values of n and M given in the table are never exactly those stated. They diverge from them by up to four or five in the third place of decimals for low n and less for large n .

We would not expect (20) to hold if S were to increase much beyond 3, since $2\sqrt{S/(1+S)}$ is starting to diverge too far from one for $1-2\sqrt{S/(1+S)}$ to be considered small compared with one.

4. COMPARISON WITH ARMITAGE'S TEST

Armitage's test has a given power for any fixed π_1/π_3 , regardless of the value of π_2 . Our test has a given power for a fixed π_1/π_3 and a given value of π_2 . If our test is designed for a given value of π_2 and this turns out to be the correct value it can be expected that our test will require a smaller maximum number of observations for the same power. We give a numerical example of this.

In his Table 5 is given the maximum number of pairs of trials with unequal results to give a power of 0.95 and a probability 0.05 of rejection of a true null hypothesis. To make the comparison with our test we divide this maximum number by $1-\pi_2$ to take account of the discarded pairs of trials in which the processes gave equal results. We may divide this number by our value of M corresponding to this value of π_2 . The ratio then is roughly constant at 1.27. This represents, for this power and probability of rejection of a true null hypothesis, the relative total number of trials required by the two tests when the assumed π_2 is equal to the actual π_2 .

Suppose, however, the true value of π_2 differs from that assumed in the construction of our test. Then if the true value is bigger than that assumed the power will decrease, if it is smaller it will increase. We can make an estimate of how big the change in power is likely to be when π_2 varies, by recalling the approximate result (20). If this holds we can take any values of M and n and map the power function of the test by varying π_2 and S . We have done this and present the result in Fig. 1. From this figure we can get approximate values of M and n for a test in which the power takes any value between 0.50 and 0.99 and the probability of acceptance of a true null hypothesis lies between 0.917 and 0.993.

Fig. 1 shows that there may be a marked drop in power if π_2 is much bigger than was assumed. This is a serious weakness of the test if there is no knowledge of the likely value. Then it will be safer to use Armitage's test.

However, we consider one case in which the lack of knowledge of π_2 may prove no bar to the use of our test. Suppose we are comparing a standard process the probability of the event happening is P_1 and for the new process the probability is P_2 . By reversing if necessary the definition of the happening of the event determined by a process we may suppose P_1 is greater than one-half. An improvement on the control process may correspond to a value of P_2 either (a) greater than P_1 ; or (b) less than P_1 .

EFFECT OF NON-NORMALITY ON THE POWER OF THE ANALYSIS OF VARIANCE TEST

By A. B. L. SRIVASTAVA

Indian Institute of Technology, Kharagpur

1. INTRODUCTION

A number of studies have been made of the effect of non-normality on the test functions used for the analysis of variance. The effect on Type I error, i.e. on the distribution under the null hypothesis, was studied by Pearson (1931), Geary (1947) and Gayen (1950), and that on Type II error, i.e. on the power function, has been considered in a study by David & Johnson (1951). Pearson showed that while both 'between-groups' and 'within-groups' mean squares, in the case of non-normal variation, provide unbiased estimates of population variance, they are no longer independently distributed; in fact, their variances and covariance contain a term in λ_4 .^{*} However, he reached the conclusion that non-normality would not have a serious effect on the distribution of their ratio w , in large samples. Considering the effect of kurtosis only, Geary (1947) gave an approximate formula for the probability correction for w , based on the large sample assumption. Gayen (1950) derived the distribution of w for non-normal populations specified by the first four terms of the Edgeworth series. He found it to consist of corrective terms in λ_4 and λ_3^2 in addition to the normal theory probability density function of w . Assuming that the observation x_{ij} contains an error term e_{ij} which has a distribution of any form whatever, with all the cumulants existing and varying from group to group, David & Johnson (1951) obtained the moments of the distribution of a certain function of the observations which makes the study of power possible in the non-normal case. They did not, however, make direct use of the distribution of a non-central variance-ratio as is done in the case of the 'normal theory' power function.

Tang (1938) and Patnaik (1949) studied the power of the normal theory χ^2 - and F -tests, by deriving the corresponding non-central distributions. The non-central distribution arises when the hypothesis of equal means is not true. The effect of non-normality on the power of the analysis of variance test can likewise be studied by investigating the non-central distribution of the variance-ratio. This has been derived in the present paper, on the assumption that the distribution of the error term is represented by the first four terms of the Edgeworth series. In addition to the normal theory power (Tang, 1938), the corrective terms in λ_3 , λ_4 and λ_3^2 have been determined. In deriving conclusions we have kept in view the fact that only the values of λ_3 and λ_4 within certain limits (Barton & Dennis, 1952) can be permitted if the Edgeworth series is to represent a positive definite and unimodal frequency function.

We have considered here the simple case of k groups of n observations. Gayen (1950) obtained the distribution of the variance-ratio assuming unequal number of observations in different groups, but we have avoided this assumption as it would necessitate introduction of more parameters of non-centrality and would make the results more complicated. The results obtained here enable one to calculate the effect of non-normality on the probability of the error of second kind, and hence on the power, when the standard F -table significance levels are used. The difficulty of having a systematic tabulation of the derived expression for

* Here and below we use λ_3 and λ_4 to denote the standardized third and fourth cumulants of the variables.

the power arises from the fact that it contains confluent hypergeometric functions, which do not appear to have been tabulated so far in detail.

2. DISTRIBUTION OF THE NON-CENTRAL VARIANCE-RATIO

Let us consider k groups of observations x_{ij} ($i = 1, 2, \dots, n; j = 1, 2, \dots, k$), and suppose that

$$x_{ij} = A + B_j + e_{ij}, \quad (1)$$

where A is the grand mean, B_j the deviation of the j th group from the grand mean so that $\sum B_j = 0$ and e_{ij} represents the random residual distributed non-normally with mean zero, variance σ^2 and standardized third and fourth cumulants $\lambda_3 (= \sqrt{\beta_1})$ and $\lambda_4 (= \beta_2 - 3)$. Assuming the higher order cumulants to be zero, we can express the distribution of the standardized variable $z_{ij} (= e_{ij}/\sigma)$, by the first four terms of the Edgeworth series as

$$f(z) = \phi(z) - \frac{\lambda_3}{6} \phi^{(3)}(z) + \frac{\lambda_4}{24} \phi^{(4)}(z) + \frac{\lambda_3^2}{72} \phi^{(6)}(z), \quad (2)$$

where $\phi(z)$ is the standard normal function, and $\phi^{(v)}(z)$ its v th derivative.

Now 'between-groups' and 'within-groups' sums of squares can be written as

$$X = \frac{n}{\sigma^2} \sum_{j=1}^k (\bar{x}_j - \bar{x})^2 = n \sum_{j=1}^k (\bar{z}_j - \bar{z} + \delta_j)^2, \quad \text{where } \delta_j = B_j/\sigma \quad (3)$$

$$Y = \frac{1}{\sigma^2} \sum_{j=1}^k \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 = \sum_{j=1}^k \sum_{i=1}^n (z_{ij} - \bar{z}_j)^2. \quad (4)$$

In the analysis of variance, as is known, we test the hypothesis H_0 that $\delta_j = 0$ for all j , and use the distribution of

$$w = [(N-k)X]/[(k-1)Y]$$

(with $\delta_j = 0$ for all j) in order to determine the point w_0 such that the probability $P(w \geq w_0)$ has a predetermined value α . We reject the hypothesis H_0 only if $w > w_0$, and accept it otherwise. If H_0 is not true, we can determine from the distribution of w (with δ_j 's not all zero), the probability of accepting H_0 , that is, the probability of the error of second kind.

Now, in order to obtain the power $\beta = 1 - P_{II}$, we first derive the non-central distribution of w . Let us put

$$S_{1j} = \sum_{i=1}^n z_{ij} + n\delta_j, \quad S_{2j} = \sum_{i=1}^n (z_{ij} - \bar{z}_j)^2 \quad (5)$$

$$X_1 = \sum_{j=1}^k S_{1j}, \quad X_2 = \frac{1}{n} \sum_{j=1}^k S_{2j}, \quad (6)$$

so that $X = X_2 - X_1^2/N$, where $N = nk$.

Using the joint distribution of S_{1j} and S_{2j} which is obtained by replacing S_{1j} by $S_{1j} - n\delta_j$ in the result (2.11) of Gayen (1949), the joint characteristic function of X_1, X_2 and Y has been derived as

$$\begin{aligned} \phi(t_1, t_2, u) = & e^{\lambda \tau_{22}} \phi_0(t_1, t_2, u) + \frac{\exp \left\{ -\frac{1}{2} (N t_1^2 - 2 i t_2 \lambda) / (1 - 2 i t_2) \right\} \tau_{22}^2 \left[\frac{2 \lambda_3}{3} \{ 3 \lambda \tau_{12} + 2 \mu \tau_{22} \} \right.}{(1 - 2 i t_2)^{\frac{1}{2} k} (1 - 2 i u)^{\frac{1}{2} (N-k)} + 4 \mu \tau_{12} \tau_{22} + 2 \nu \tau_{22}^2} \\ & + \frac{\lambda_4}{3} \left\{ 3 \lambda \left(\tau_{12}^2 - 1 + \frac{k}{N(1 - 2 i t_2)} + \frac{N-k}{N(1 - 2 i u)} \right) + \frac{3}{N(1 - 2 i t_2)} [(k+3) \tau_{12}^2 - (N+3k)/N] \right. \\ & + \frac{\lambda_3^2}{9} \left\{ 3 \lambda \left([N \tau_{12}^4 - 3(N+3) \tau_{12}^2 + 6] + \frac{3(N-k)}{(1 - 2 i t_2)} \right) + \frac{3(N-k)}{N(1 - 2 i u)} + \frac{3(N-k)}{N(1 - 2 i u)^2} \right\} \\ & + \frac{6k}{N(1 - 2 i t_2)^2} + \frac{3(N-k)}{N(1 - 2 i u)} (N \tau_{12}^2 - 3) + \frac{3(N-k)}{N(1 - 2 i u)(1 - 2 i t_2)} + \frac{3(N-k)}{N(1 - 2 i u)^2} \\ & + 2 \mu \tau_{12} \tau_{22} \left([N \tau_{12}^2 - 3(N+6)] + \frac{3(k+6)}{(1 - 2 i t_2)} + \frac{3(N-k)}{(1 - 2 i u)} \right) \\ & \left. + 18 \tau_{22}^2 (\lambda^2 \tau_{12}^2 + 2 \nu \tau_{22}) + 24 \lambda \mu \tau_{12} \tau_{22}^3 + 8 \mu^2 \tau_{22}^4 \right\}. \end{aligned} \quad (7)$$

where

$$\tau_{12} = \frac{it_1}{(1-2it_2)}, \quad \tau_{22} = \frac{it_2}{(1-2it_2)}; \quad (8)$$

$$\lambda = n \sum_1^k \delta_j^2, \quad \mu = n \sum_1^k \delta_j^3 \quad \text{and} \quad \nu = n \sum_1^k \delta_j^{4*} \quad (9)$$

and where $\phi_0(t_1, t_2, u)$ denotes Gayen's (1950) result (2.6)[†], in which we have taken $k'^2 = k^2$.

Obtaining the joint frequency function of X_1, X_2 and Y from (7) by Fourier's inversion theorem, and integrating out X_1 , we find the joint frequency function of X and Y can be written in the form

$$g(\cdot, X, Y) = g_0(\cdot, X, Y) + \lambda_3 g_{\lambda_3}(\cdot, X, Y) + \lambda_4 g_{\lambda_4}(\cdot, X, Y) + \lambda_3^2 g_{\lambda_3^2}(\cdot, X, Y), \quad (10)$$

$$\text{the typical term in it being} \quad \phi(Y, r_1) \psi(X, r_2), \quad (11)$$

$$\text{where} \quad \phi(Y, r_1) = \frac{Y^{\frac{1}{2}r_1-1} e^{-\frac{1}{2}Y}}{2^{\frac{1}{2}r_1} \Gamma(\frac{1}{2}r_1)} \quad (12)$$

$$\text{and} \quad \psi(X, r_2) = \frac{e^{-\frac{1}{2}\lambda} e^{-\frac{1}{2}X}}{2^{\frac{1}{2}r_2}} \sum_{j=0}^{\infty} \frac{X^{\frac{1}{2}r_2+j-1} \lambda^j}{2^{2j} \Gamma(\frac{1}{2}r_2+j) j!}. \quad (13)$$

We can now deduce the frequency function of $w = [(N-k)X]/[(k-1)Y]$ from (10) in the form

$$g(w) = g_0(w) + \lambda_3 g_{\lambda_3}(w) + \lambda_4 g_{\lambda_4}(w) + \lambda_3^2 g_{\lambda_3^2}(w). \quad (14)$$

Using $p(w; r_1, r_2)$ to denote

$$\sum_{j=0}^{\infty} \frac{e^{-\frac{1}{2}\lambda} (\frac{1}{2}\lambda)^j}{j! B(\frac{1}{2}r_1+j, \frac{1}{2}r_2)} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{1}{2}r_1+j} \frac{w^{\frac{1}{2}r_1+j-1}}{\left(1 + \frac{\nu_1}{\nu_2} w\right)^{\frac{1}{2}(r_1+r_2)+j}}, \quad (15)$$

where $\nu_1 = k-1, \nu_2 = N-k$ are the degrees of freedom of X and Y , we find

$$g_0(w) = p(w; \nu_1, \nu_2), \quad (16)$$

$$g_{\lambda_3}(w) = \frac{1}{6} \mu p_3(w; \nu_1, \nu_2) \quad (17)$$

$$\begin{aligned} g_{\lambda_4}(w) = & \frac{1}{24N} [3\{(\nu_1 + \nu_2)^2 p(w; \nu_1, \nu_2) - 2\nu_1(\nu_1 + \nu_2) p(w; \nu_1 + 2, \nu_2) \\ & + \nu_1^2 p(w; \nu_1 + 4, \nu_2) - 2\nu_2(\nu_1 + \nu_2) p(w; \nu_1, \nu_2 + 2) + 2\nu_1\nu_2 p(w; \nu_1 + 2, \nu_2 + 2) \\ & + \nu_2^2 p(w; \nu_1, \nu_2 + 4)\} + 6\lambda\{-(\nu_1 + \nu_2) p_2(w; \nu_1, \nu_2) + \nu_1 p_2(w; \nu_1 + 2, \nu_2) \\ & + \nu_2 p_2(w; \nu_1, \nu_2 + 2)\} + N\nu p_4(w; \nu_1, \nu_2)], \end{aligned} \quad (18)$$

$$\begin{aligned} g_{\lambda_3^2}(w) = & \frac{1}{72N} [6\{-(\nu_1 + \nu_2)(\nu_1 + \nu_2 - 1) p(w; \nu_1, \nu_2) + 3\nu_1(\nu_1 + \nu_2 - 1) p(w; \nu_1 + 2, \nu_2) \\ & - 3\nu_1(\nu_1 - 1) p(w; \nu_1 + 4, \nu_2) + \nu_1(\nu_1 - 1) p(w; \nu_1 + 6, \nu_2) \\ & + 3\nu_2(\nu_1 + \nu_2 - 1) p(w; \nu_1, \nu_2 + 2) - 6\nu_1\nu_2 p(w; \nu_1 + 2, \nu_2 + 2) \\ & - 3\nu_2(\nu_2 - 1) p(w; \nu_1, \nu_2 + 4) + 3\nu_1\nu_2 p(w; \nu_1 + 2, \nu_2 + 4) \\ & + \nu_2(\nu_2 - \nu_1 - 1) p(w; \nu_1, \nu_2 + 6)\} + 18\lambda\{(\nu_1 + \nu_2 - 1) p_2(w; \nu_1, \nu_2) \\ & - 2(\nu_1 - 1) p_2(w; \nu_1 + 2, \nu_2) + (\nu_1 - 1) p_2(w; \nu_1 + 4, \nu_2) - 2\nu_2 p_2(w; \nu_1, \nu_2 + 2) \\ & + \nu_2 p_2(w; \nu_1, \nu_2 + 4)\} + 9(\lambda^2 - N\nu)\{p_4(w; \nu_1, \nu_2) - p_4(w; \nu_1 + 2, \nu_2)\} \\ & + N\mu^2 p_6(w; \nu_1, \nu_2)], \end{aligned} \quad (19)$$

* The conventional λ , or non-central parameter, used here should not be confused with the standardized cumulants, λ_3 and λ_4 , nor should ν be confused with the degrees of freedom, ν_1 and ν_2 , defined below.

† In the first line of formula (2.6) of Gayen (1950), the factor $(1-2it_2)^{\frac{1}{2}}$ in the denominator should read as $(1-2it_2)^{\frac{1}{2}k}$.

where

$$p_r(w; r_1, r_2) = p(w; r_1 + 2r, r_2) - rp(w; r_1 + 2r - 2, r_2) \\ + \frac{r(r-1)}{2!} p(w; r_1 + 2r - 4, r_2) - \dots + (-1)^r p(w; r_1, r_2). \quad (20)$$

3. POWER FUNCTION

We are now in a position to evaluate the probability of the error of second kind P_{II} by integrating $g(w)$ over the region of acceptance. If w_0 be the point which determines the critical region, we shall have

$$P_{II}(w_0) = \int_0^{w_0} g(w) dw = P_0(w_0) + \lambda_3 P_{\lambda_3}(w_0) + \lambda_4 P_{\lambda_4}(w_0) + \lambda_5^2 P_{\lambda_5^2}(w_0). \quad (21)$$

In this, $P_0(w_0)$ as obtained from (16), is given by

$$P_0(w_0) = \sum_{j=0}^{\infty} \frac{e^{-\frac{1}{2}\lambda} (\frac{1}{2}\lambda)^j}{j!} I_{u_0}(\frac{1}{2}\nu_1 + j, \frac{1}{2}\nu_2), \quad (22)$$

where $u_0 = (k-1)w_0 / [(N-k) + (k-1)w_0]$ and I_{u_0} denotes the incomplete β -function. This is the normal-theory expression for $P_{II}(w_0)$ and the methods of its evaluation have been discussed by Tang (1938) and Patnaik (1949). Here the expressions for $P_{\lambda_3}(w_0)$, $P_{\lambda_4}(w_0)$ and $P_{\lambda_5^2}(w_0)$ as obtained by integrating (17), (18) and (19), respectively, are expressible in terms of the confluent hypergeometric functions

$$F(a, b, x) = \sum_{j=0}^{\infty} \frac{\Gamma(a+j) \Gamma(b) x^j}{\Gamma(b+j) \Gamma(a) j!}. \quad (23)$$

Writing

$$F_{(0)}(a, b, \frac{1}{2}\lambda u_0) = \frac{\Gamma(a)}{\Gamma(a-b+1) \Gamma(b)} F(a, b, \frac{1}{2}\lambda u_0),$$

and

$$F_{(r)}(a, b, \frac{1}{2}\lambda u_0) = F_{(0)}(a, b, \frac{1}{2}\lambda u_0) - r C_1 u_0 F_{(0)}(a+1, b+1, \frac{1}{2}\lambda u_0) \\ + \dots + (-1)^r u_0^r F_{(0)}(a+r, b+r, \frac{1}{2}\lambda u_0), \quad (24)$$

$$F_{(r)'}(a, b, \frac{1}{2}\lambda u_0) = F_{(0)}(a, b, \frac{1}{2}\lambda u_0) - r C_1 (1-u_0) F_{(0)}(a+1, b, \frac{1}{2}\lambda u_0) \\ + \dots + (-1)^r (1-u_0)^r F_{(0)}(a+r, b, \frac{1}{2}\lambda u_0), \quad (25)$$

we get

$$P_{\lambda_3}(w_0) = -\frac{1}{6}\mu e^{-\frac{1}{2}\lambda} u_0^{\frac{1}{2}\nu_1} (1-u_0)^{\frac{1}{2}\nu_2} F_{(2)}\left(\frac{\nu_1+\nu_2}{2}, \frac{\nu_1+2}{2}, \frac{1}{2}\lambda u_0\right), \\ P_{\lambda_4}(w_0) = \frac{1}{8N} e^{-\frac{1}{2}\lambda} u_0^{\frac{1}{2}\nu_1} (1-u_0)^{\frac{1}{2}\nu_2} \left[\nu_1^2 F_{(1)}\left(\frac{\nu_1+\nu_2}{2}, \frac{\nu_1+2}{2}, \frac{1}{2}\lambda u_0\right) \right. \\ \left. - \nu_2^2 F_{(1)'}\left(\frac{\nu_1+\nu_2}{2}, \frac{\nu_1}{2}, \frac{1}{2}\lambda u_0\right) - 2\nu_1\nu_2 F_{(1)}\left(\frac{\nu_1+\nu_2}{2}, \frac{\nu_1}{2}, \frac{1}{2}\lambda u_0\right) \right. \\ \left. - 2\lambda \left\{ \nu_1 F_{(2)}\left(\frac{\nu_1+\nu_2}{2}, \frac{\nu_1+2}{2}, \frac{1}{2}\lambda u_0\right) - \nu_2 F_{(2)}\left(\frac{\nu_1+\nu_2}{2}, \frac{\nu_1}{2}, \frac{1}{2}\lambda u_0\right) \right\} \right. \\ \left. + \frac{N}{3} \nu F_{(3)}\left(\frac{\nu_1+\nu_2}{2}, \frac{\nu_1+2}{2}, \frac{1}{2}\lambda u_0\right) \right], \quad (26)$$

$$P_{\lambda_5^2}(w_0) = \frac{1}{12N} e^{-\frac{1}{2}\lambda} u_0^{\frac{1}{2}\nu_1} (1-u_0)^{\frac{1}{2}\nu_2} \left[-\nu_1(\nu_1-1) F_{(2)}\left(\frac{\nu_1+\nu_2}{2}, \frac{\nu_1+2}{2}, \frac{1}{2}\lambda u_0\right) \right. \\ \left. + \nu_2(\nu_2-1) F_{(2)}\left(\frac{\nu_1+\nu_2}{2}, \frac{\nu_1}{2}, \frac{1}{2}\lambda u_0\right) - \nu_1\nu_2 \left\{ 3F_{(2)'}\left(\frac{\nu_1+\nu_2}{2}, \frac{\nu_1+2}{2}, \frac{1}{2}\lambda u_0\right) \right. \right. \\ \left. \left. + F_{(2)'}\left(\frac{\nu_1+\nu_2}{2}, \frac{\nu_1}{2}, \frac{1}{2}\lambda u_0\right) \right\} + 3\lambda \left\{ (\nu_1-1) F_{(3)}\left(\frac{\nu_1+\nu_2}{2}, \frac{\nu_1+2}{2}, \frac{1}{2}\lambda u_0\right) \right. \right. \\ \left. \left. + \nu_2 F_{(3)}\left(\frac{\nu_1+\nu_2}{2}, \frac{\nu_1-2}{2}, \frac{1}{2}\lambda u_0\right) \right\} + \frac{3}{2}(\lambda^2 - N\nu) F_{(4)}\left(\frac{\nu_1+\nu_2}{2}, \frac{\nu_1+2}{2}, \frac{1}{2}\lambda u_0\right) \right. \\ \left. + \frac{N}{6} \mu^2 F_{(5)}\left(\frac{\nu_1+\nu_2}{2}, \frac{\nu_1+2}{2}, \frac{1}{2}\lambda u_0\right) \right]. \quad (27)$$

If ν_2 is an even integer, the above expressions can be reduced to the much simpler form of finite series by using the formula

$$F(a+r, a, x) = e^x F(-r, a, -x) = e^x \sum_{i=0}^r \frac{B(a, r+1)}{B(a+i, r+1-i)} \frac{x^i}{i!}. \quad (28)$$

For the normal-theory term, $P_0(w_0)$, such an expression in the form of a finite series when ν_2 is even, has already been given by Tang (1938). It is found that by using (28) the $F_{(r)}$ and $F_{(r)'}'$ functions occurring in the corrective terms yield expressions which are not difficult to evaluate for ν_2 as large as 20. For example, in order to evaluate $P_{\lambda_3}(w_0)$ we have to calculate the value of $F_{(2)}\left(\frac{1}{2}(\nu_1 + \nu_2), \frac{1}{2}(\nu_1 + 2), \frac{1}{2}\lambda u_0\right)$ which can be easily done by using the formula

$$F_{(2)}\left(\frac{\nu_1 + \nu_2}{2}, \frac{\nu_1 + 2}{2}, \frac{1}{2}\lambda u_0\right) = e^{\frac{1}{2}\lambda u_0} \sum_{i=0}^{\frac{\nu_2}{2}-1} \left[\left(\frac{1}{2}(\nu_1 + \nu_2 + 2) \right) u_0^2 - 2 \left(\frac{1}{2}(\nu_1 + \nu_2) \right) u_0 + \left(\frac{1}{2}(\nu_1 + \nu_2 - 2) \right) \right] \frac{\left(\frac{1}{2}\lambda u_0 \right)^i}{i!},$$

where

$$\binom{a}{b} = \frac{\Gamma(a+1)}{\Gamma(b+1) \Gamma(a-b+1)}.$$

Similar expressions can be worked out for $F_{(1)}$, $F_{(1)'}$, $F_{(2)'}$, $F_{(3)}$, etc., which occur in the equations (26) and (27), in order to simplify the computation of $P_{\lambda_4}(w_0)$ and $P_{\lambda_3^2}(w_0)$ when ν_2 is even.

4. NUMERICAL EVALUATION OF RESULTS AND DISCUSSION

We shall now consider numerical examples to illustrate the nature of the effects of non-normality on the power. It should be mentioned here that the formulae given above are useful only in determining the effects of such non-normality as is not of very serious type. However, a considerable range of values of λ_3 and λ_4 can be covered within the limits given by Barton & Dennis (1952) (see § 1 above).

The values of $P_{\lambda_3}(w_0)$, $P_{\lambda_4}(w_0)$ and $P_{\lambda_3^2}(w_0)$ have been calculated for $\nu_1 = 4$, $\nu_2 = 10$ and $\nu_1 = 4$, $\nu_2 = 20$ when the critical region is determined by the 'normal theory' upper 5% values of F and the alternatives are given by $\phi = \sqrt{(\lambda/k)} = 1.0, 1.5, 2.0, 2.5$, etc. This critical region is, of course, based on an erroneous F -value; the actual probability of Type I error, however, can be obtained by adding to 0.05 the corrections due to non-normality given by Gayen (1950)*. In the examples considered, since ν_2 is even, it has been possible to use the finite series expansions for calculation. The values are given in Table 1.

It is first necessary to examine what values μ and ν can take when $\lambda = k\phi^2$ has some fixed positive value. We have seen that the null-hypothesis H_0 assumes that δ_j 's, the deviations of the group means from the grand mean, are all zero. If differences between groups exist, the δ_j 's will not be all zero and their set of values will define an alternative to H_0 . In practice, the true values of group differences are not known, and one would depend on the parameters λ , μ and ν determined by δ_j 's to define an alternative hypothesis. As indicated by Tang (1938) in an example, two extreme forms of the set of δ_j 's under the restriction $\sum_1^k \delta_j = 0$ will be as follows:

- the k groups form two distinct equal sets (with a single group midway between if k is odd), the effects in one set being equal and opposite in sign to those in the other;
- of the k groups, $k-1$ are quite similar and only one is divergent, with its effect counterbalancing the rest.

* There are a few slips in Gayen's (1950, pp. 242) formulae (2.30) to (2.33 bis). There should be a minus sign before λ_4 in (2.30) instead of the plus sign and in all the formulae from (2.31) to (2.33 bis) x_0 should be replaced by $1 - x_0$.

For a fixed λ , in the case (a) the value of μ will be zero and that of ν will be minimum ($= \lambda^2/N$ or $k\lambda^2/[N(k-1)]$ according as k is even or odd). In the case (b), μ will have an extreme positive or negative value ($= \pm(k-2)\lambda^{\frac{3}{2}}/[N(k-1)]^{\frac{1}{2}}$) and ν will be maximum ($= (k^2-3k+3)\lambda^2/[N(k-1)]$). In practice the situation may be anywhere between these two extreme cases.

Table 1. Showing the values of the corrective functions for $\alpha = 0.05$ required in equation (21) when (i) $\nu_1 = 4, \nu_2 = 10$ and (ii) $\nu_1 = 4, \nu_2 = 20$

D.F.	ϕ	$P_0(w_0)$	$P_{\lambda_3}(w_0)$	$P_{\lambda_4}(w_0)$	$P_{\lambda_3^2}(w_0)$
$\nu_1 = 4$ $\nu_2 = 10$	0.0	0.950	0	0.00376	-0.00429
	1.0	0.738	0.0010750 μ	-0.01273 - 0.0000744 ν	0.00480 - 0.0000213 ν + 0.00000604 μ^2
	1.5	0.449	0.0002396 μ	-0.01916 - 0.0000489 ν	0.00300 + 0.0000422 ν - 0.00000007 μ^2
	2.0	0.179	-0.0001670 μ	-0.00156 - 0.0000045 ν	-0.00390 + 0.0000157 ν - 0.00000067 μ^2
	2.5	0.043	-0.0001089 μ	0.00944 + 0.0000047 ν	-0.00174 - 0.0000008 ν - 0.00000008 μ^2
	3.0	0.006	-0.0000239 μ	0.00495 + 0.0000015 ν	0.00085 - 0.0000010 ν + 0.00000002 μ^2
$\nu_1 = 4$ $\nu_2 = 20$	0.0	0.950	0	0.00241	-0.00096
	1.0	0.681	0.0018003 μ	-0.01122 - 0.0001695 ν	0.00273 - 0.0000038 ν + 0.00001832 μ^2
	1.5	0.341	0.0000834 μ	-0.00929 - 0.0000818 ν	-0.00159 + 0.0001133 ν - 0.00000190 μ^2
	2.0	0.093	-0.0003425 μ	0.00050 + 0.0000083 ν	-0.00182 + 0.0000169 ν - 0.00000148 μ^2
	2.5	0.012	-0.0000978 μ	0.00458 + 0.0000074 ν	0.00084 - 0.0000083 ν + 0.00000006 μ^2

N.B. The parameter λ, μ and ν have been defined in equations (9).

In Table 2, the values of $P_{\lambda_3}(w_0), P_{\lambda_4}(w_0)$ and $P_{\lambda_3^2}(w_0)$ (of Table 1) are given for the following cases: (a) $\mu = 0, \nu$ maximum; (b₁) μ maximum positive, ν minimum; and (b₂) μ maximum negative, ν minimum. The sign of μ , which distinguishes case (b₂) from (b₁), only affects $P_{\lambda_3}(w_0)$. Where there are two signs with a value for case (b), the upper sign relates to (b₁). The values tabled are expected to be correct to five places of decimals.

In the examples, the effects of λ_3, λ_4 and λ_3^2 are observed generally in the third place of decimals but sometimes in the second place. The effect of λ_4 is, in general, higher than that of λ_3 or λ_3^2 . Also it appears that the values of $P_{\lambda_4}(w_0)$, as well as of $P_{\lambda_3^2}(w_0)$, are not much different in the cases (a) and (b).

We must now compute the power of the test in the case of non-normal populations with given values of λ_3 and λ_4 . In Table 3, such comparison is given for the examples considered above. Here the two entries before any value of ϕ in a column with $\lambda_3 = 0$ correspond to the cases (a) and (b), respectively, and the three entries for $\lambda_3 \neq 0$, correspond to the cases (a), (b₁) and (b₂). It is to be noted that the values for the cases (b₁) and (b₂) will interchange as λ_3 changes sign.

Table 2. Giving the values of $P_{\lambda_3}(w_0)$, $P_{\lambda_4}(w_0)$ and $P_{\lambda_3^2}(w_0)$ (of Table 1) in the cases (a), (b₁) and (b₂)

ϕ	Case	$\nu_1 = 4, \nu_2 = 10$			$\nu_1 = 4, \nu_2 = 20$		
		$P_{\lambda_3}(w_0)$	$P_{\lambda_4}(w_0)$	$P_{\lambda_3^2}(w_0)$	$P_{\lambda_3}(w_0)$	$P_{\lambda_4}(w_0)$	$P_{\lambda_3^2}(w_0)$
0.0		0	0.00376	-0.00429	0	0.00241	-0.00096
1.0	(a)	0	-0.01289	0.00476	0	-0.01143	0.00273
	(b)	± 0.00465	-0.01313	0.00479	± 0.00604	-0.01177	0.00293
1.5	(a)	0	-0.01968	-0.00345	0	-0.00981	-0.00087
	(b)	± 0.00350	-0.02050	-0.00415	± 0.00094	-0.01064	0.00003
2.0	(a)	0	-0.00171	-0.00338	0	-0.00067	-0.00148
	(b)	∓ 0.00579	-0.00195	-0.00334	∓ 0.00919	0.00093	-0.00201
2.5	(a)	0	0.00982	-0.00181	0	0.00494	0.00058
	(b)	∓ 0.00737	0.01043	-0.00228	∓ 0.00502	0.00552	0.00032
3.0	(a)	0	0.00520	0.00068	—	—	—
	(b)	∓ 0.00279	0.00561	0.00068	—	—	—

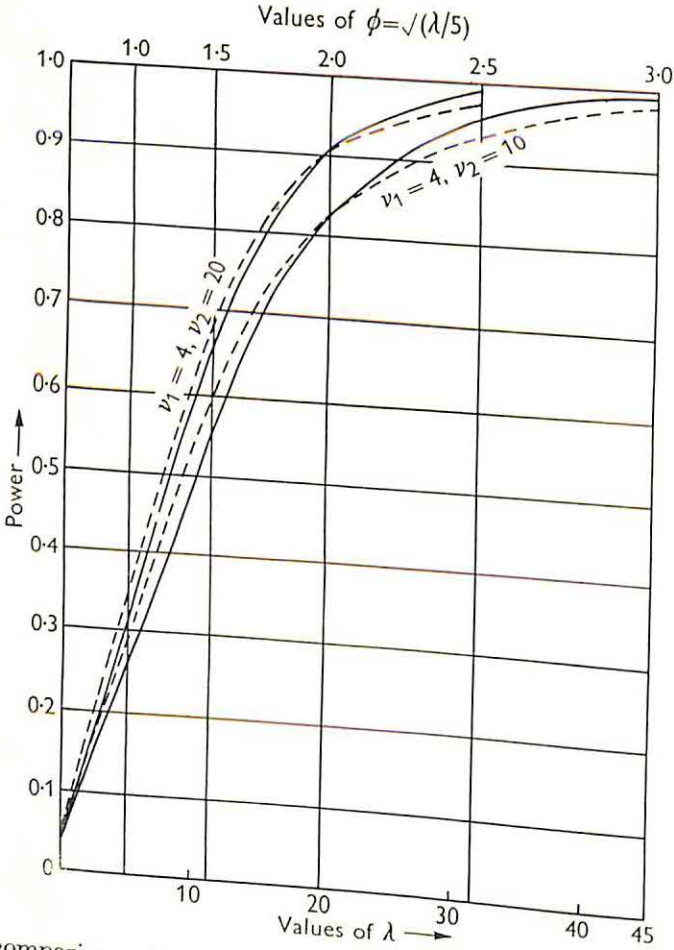


Fig. 1. Showing comparison of powers. —, normal; - - -, non-normal ($\lambda_3 = 0, \lambda_4 = 2.4$).

It is again seen that when $\lambda_3 = 0$ there is practically no difference between the values of power in the situations (a) and (b). When $\lambda_3 \neq 0$, in the three situations (a), (b₁) and (b₂) that arise, the powers differ to some extent though not very significantly from the practical point of view. The situation (a), in this case, gives rise to values which are generally intermediate between those of (b₁) and (b₂).

Table 3. *Showing comparison of powers of the F-test in normal and non-normal cases when using the normal theory with 5% significance level*

ϕ	Symmetrical distributions, $\lambda_3 = 0.0$						Skew distributions, $\lambda_3 = 0.5$				$\lambda_3 = 0.7$
	$\lambda_4 = -1.0$	$\lambda_4 = 0.0$	$\lambda_4 = 0.5$	$\lambda_4 = 1.0$	$\lambda_4 = 2.0$	$\lambda_4 = 2.4$	$\lambda_4 = -1.0$	$\lambda_4 = 0.0$	$\lambda_4 = 0.5$	$\lambda_4 = 2.0$	$\lambda_4 = 2.4$
Case (i) $\nu_1 = 4, \nu_2 = 10$											
0.0	0.054	0.050	0.048	0.046	0.042	0.041	0.055	0.051	0.049	0.044	0.043
1.0	0.249 .249	0.262 —	0.268 .269	0.275 .275	0.288 .288	0.290 .294	0.248 .245 .250	0.261 .258 .263	0.267 .265 .270	0.287 .285 .289	0.291 .288 .294
1.5	0.531 .531	0.551 —	0.561 .561	0.571 .572	0.590 .592	0.598 .600	0.532 .530 .533	0.552 .550 .554	0.562 .566 .564	0.591 .591 .595	0.600 .600 .605
2.0	0.819 .819	0.821 —	0.822 .822	0.823 .823	0.824 .825	0.825 .826	0.820 .823 .817	0.822 .825 .819	0.823 .826 .820	0.825 .829 .823	0.827 .831 .823
2.5	0.967 .967	0.957 —	0.952 .952	0.947 .947	0.937 .936	0.933 .932	0.967 .972 .964	0.957 .961 .954	0.953 .956 .949	0.938 .940 .933	0.934 .938 .928
3.0	0.999 —	0.994 —	0.991 .991	0.989 .988	0.984 .983	0.982 .981	0.999 .999 .998	0.994 .995 .992	0.991 .992 .990	0.983 .984 .981	0.981 .982 .978
Case (ii) $\nu_1 = 4, \nu_2 = 20$											
0.0	0.052	0.050	0.049	0.048	0.045	0.044	0.053	0.050	0.049	0.045	0.045
1.0	0.308 .307	0.319 —	0.325 .325	0.330 .331	0.342 .343	0.346 .347	0.307 .303 .310	0.318 .315 .321	0.324 .321 .327	0.341 .339 .345	0.345 .342 .350
1.5	0.649 .648	0.659 —	0.664 .664	0.669 .670	0.679 .680	0.683 .685	0.649 .648 .649	0.659 .658 .659	0.664 .664 .665	0.679 .680 .681	0.683 .684 .685
2.0	0.908 .908	0.907 —	0.907 .907	0.906 .906	0.906 .905	0.905 .905	0.908 .913 .904	0.907 .912 .903	0.907 .912 .902	0.906 .910 .901	0.906 .912 .899
2.5	0.993 .994	0.988 —	0.986 .985	0.983 .982	0.978 .977	0.976 .975	0.993 .996 .991	0.988 .990 .985	0.985 .988 .983	0.978 .979 .974	0.976 .978 .971

Power curves relating to the symmetric leptokurtic population with $\lambda_4 = 2.4$ ($\beta_2 = 5.4$) have been drawn in Fig. 1 along with the 'normal theory' power curves for the cases $\nu_1 = 4, \nu_2 = 10$ and $\nu_1 = 4, \nu_2 = 20$. This diagram illustrates the nature and magnitude of the effect of such a departure from normality in the distribution of the random residuals. Power curves in the case of the other non-normal populations considered in Table 3 would differ from normal theory to a less extent.

5. CONCLUSION

Obtaining the expressions for the power function of the analysis of variance test for a one-way classification in the case where the sampled population is represented by the first four terms of an Edgeworth series, we have considered certain numerical examples to see what is the nature of effects of non-normality on the power. The populations considered are only moderately non-normal, as the terms higher than the fourth in the Edgeworth series are assumed to be zero. Also the values of λ_3 and λ_4 have not been allowed to exceed the Barton & Dennis (1952) limits, ensuring thereby that the frequency function represented is positive definite and unimodal.

It is found that the effect of skewness is not much on the power of the analysis of variance test, at least when the values of λ_3 are confined to Barton & Dennis limits. In practice, however, higher values of λ_3 may occur and then the effect of skewness may be somewhat larger and comparable to that of kurtosis which is, in general, high. The presence of a fair degree of kurtosis, as is not uncommon in practice, leads to a noticeable change in the power curve particularly in the case of small samples. But a small departure from normality in respect of kurtosis (say, of the order $\lambda_4 = \pm 0.5$) again does not cause any significant deviation in the power. When the population is leptokurtic ($\lambda_4 > 0$), the power increases in the beginning (for example, up to the point for which the power is approximately 0.8 in the case when $\nu_1 = 4, \nu_2 = 10$ (Fig. 1)), but in the region of very high power, it subsequently decreases in comparison with the normal-theory power. The reverse happens when λ_4 is negative.

The above conclusions derived from numerical results are expected to be valid in general. There is a good indication that the effect of non-normality on the power diminishes with increasing sample size, as expected. In practice the effect of kurtosis is likely to be more on the power than that of skewness. A general conclusion in this regard cannot, however, be given, for it is difficult to take into account all the ways in which the values of λ_3 and λ_4 of the distributions met in practice will vary. But on the whole it may be said that from the practical point of view, the effect of non-normality on power will not be of much consequence in the case of near-normal populations.

My thanks are due to Dr A. K. Gayen for his help and guidance in the course of preparation of this work. I am also grateful to Prof. E. S. Pearson for certain comments and suggestions leading to some improvement of the paper.

REFERENCES

- BARTON, D. E. & DENNIS, K. E. (1952). *Biometrika*, **39**, 425-7.
 DAVID, F. N. & JOHNSON, N. L. (1951). *Biometrika*, **38**, 43-57.
 GAYEN, A. K. (1949). *Biometrika*, **36**, 353-69.
 GAYEN, A. K. (1950). *Biometrika*, **37**, 236-55.
 GEARY, R. C. (1947). *Biometrika*, **34**, 209-42.
 PATNAIK, P. B. (1949). *Biometrika*, **36**, 202-32.
 PEARSON, E. S. (1931). *Biometrika*, **23**, 114-33.
 TANG, P. C. (1938). *Statist. Res. Mem.* **2**, 126-57.

THE z -TEST AND SYMMETRICALLY DISTRIBUTED RANDOM VARIABLES

By F. N. DAVID*

University College London and Statistical Laboratory, University of California, Berkeley

Some years ago in two papers (David & Johnson 1951*b*, 1952) the exact cumulants and cross-cumulants of the distributions of the sums of squares usually associated as test criteria for the general linear hypothesis were given under extremely wide assumptions. It had been the intention, at this stage, to substitute in the given formulae the expressions associated with each of the analysis of variance set-ups and thus, by taking appropriate numerical values, to obtain some idea of the effect of non-normality, with and without unequal group frequencies, on the F -test as commonly applied. The arithmetic, however, proved formidable and an algebraic alternative was sought for. This algebraic alternative has proved of use in other problems not associated with the general linear hypothesis. It is proposed therefore to summarize it here and to illustrate its use on the F -test of the one-way classification, the non-normality of which has previously been discussed by Pearson (1931), David & Johnson (1951*a*), and Gayen (1950). It is hoped to present similar formulae using the cumulants for the general linear hypothesis sums of squares in a further publication.

Assume k groups containing $n_1, n_2, \dots, n_k$ observations with

$$\sum_{t=1}^k n_t = n.$$

In commonly used notation, the one-way classification leads to the identity

$$\sum_{t=1}^k n_t (\bar{x}_t - \bar{x}_{..})^2 + \sum_{t=1}^k \sum_{i=1}^{n_t} (x_{ti} - \bar{x}_t)^2 = \sum_{t=1}^k \sum_{i=1}^{n_t} (x_{ti} - \bar{x}_{..})^2.$$

Usually the x 's are assumed each to follow the same normal distribution. We shall assume that each has the same symmetrical distribution. There are no real difficulties introduced by still further generalizing the situation and assuming for each group that x has a different asymmetrical distribution. The purpose here however is one of method and the symmetrical distribution leads to shortening of the expressions for cumulants. If

$$S_1 = \sum_{t=1}^k n_t (\bar{x}_t - \bar{x}_{..})^2 \quad \text{and} \quad S_2 = \sum_{t=1}^k \sum_{i=1}^{n_t} (x_{ti} - \bar{x}_t)^2,$$

then we may write down the cumulants and cross-cumulants of S_1 and S_2 from those already

given by David & Johnson (1951*a*). We have that

$$\mathcal{C}(S_1) = (k-1)\kappa_2 = f_1\kappa_2, \quad \mathcal{C}(S_2) = (n-k)\kappa_2 = f_2\kappa_2.$$

$$\gamma_r = \frac{\kappa_{r+2}}{\kappa_2^{r+1}},$$

We define

$$K(S_1^d S_2^e) = \frac{\kappa(S_1^d S_2^e)}{(\kappa(S_1))^d (\kappa(S_2))^e}$$

and

$$V_{abc} = \frac{n^b}{(n-k)^a k^{b+1-a-c} (k-1)^c} \sum_{t=1}^k \frac{(n_t-1)^a}{n_t^b} \left(1 - \frac{n_t}{n}\right)^c.$$

* With the partial support of the Office of Naval Research, U.S. Navy.

It will be noticed that $V_{abc} = 1$ when $n_t = n/k$, $t = 1, 2, \dots, k$. From the printed cumulants we get, putting the odd gammas zero, and rearranging

$$K(S_1^2) = \gamma_2 \frac{V_{012}}{n} + \frac{2}{f_1}, \quad K(S_1 S_2) = \gamma_2 \frac{V_{111}}{n}, \quad K(S_2^2) = \gamma_2 \frac{V_{210}}{n} + \frac{2}{f_2},$$

$$K(S_1^3) = \gamma_4 \frac{V_{023}}{n^2} + 12\gamma_2 \frac{V_{012}}{nf_1} + \frac{8}{f_1^2},$$

$$K(S_1^2 S_2) = \gamma_4 \frac{V_{122}}{n^2} + 4\gamma_2 \frac{V_{111}}{nf_1},$$

$$K(S_1 S_2^2) = \gamma_4 \frac{V_{221}}{n^2} + 4\gamma_2 \frac{V_{111}}{nf_2},$$

$$K(S_2^3) = \gamma_4 \frac{V_{320}}{n^2} + 12\gamma_2 \frac{V_{210}}{nf_2} + \frac{8}{f_2^2}.$$

$$K(S_1^4) = \gamma_6 \frac{V_{034}}{n^3} + 24\gamma_4 \frac{V_{023}}{n^2 f_1} + \gamma_2^2 \frac{32V_{023}}{n^2 f_1} + 144\gamma_2 \frac{V_{012}}{nf_1^2} + \frac{48}{f_1^3},$$

$$K(S_1^3 S_2) = \gamma_6 \frac{V_{133}}{n^3} + 12\gamma_4 \frac{V_{122}}{n^2 f_1} + \gamma_2^2 \frac{12V_{122}}{n^2 f_1} + 24\gamma_2 \frac{V_{111}}{nf_1^2},$$

$$K(S_1^2 S_2^2) = \gamma_6 \frac{V_{232}}{n^3} + 4\gamma_4 \frac{V_{111}}{nf_1 f_2} + \gamma_2^2 \left(\frac{12V_{112}}{nf_1 f_2} + \frac{4V_{211}}{nf_1 f_2} \right) + \frac{16\gamma_2 V_{111}}{nf_1 f_2},$$

$$K(S_1 S_2^3) = \gamma_6 \frac{V_{331}}{n^3} + 12\gamma_4 \frac{V_{221}}{n^2 f_2} + 4\gamma_2^2 \left(\frac{7V_{221}}{n^2 f_2} - \frac{2V_{111}}{nf_2^2} \right) + 24\gamma_2 \frac{V_{111}}{nf_2^2},$$

$$K(S_2^4) = \gamma_6 \frac{V_{430}}{n^3} + 24\gamma_4 \frac{V_{320}}{n^2 f_2} + 8\gamma_2^2 \left(\frac{6V_{320}}{n^2 f_2} - \frac{3V_{210}}{nf_2^2} + \frac{1}{f_2^3} \right) + 144\gamma_2 \frac{V_{210}}{nf_2^2} + \frac{48}{f_2^3}.$$

These expressions are all exact except that, in the first three cumulants of the fourth order, terms in γ_2^2 which were of order n^{-4} and n^{-5} have been omitted.

Fisher and Cornish in their paper dealing with the inversion of Edgeworth's expansion showed that the cumulative function of

$$z = \frac{1}{2} \log F$$

could be adequately represented in the normal case by the expansion which they obtained. Thus if the cumulants of normal z , say z_N are $\kappa_{1N}, \kappa_{2N}, \dots, \kappa_{rN}, \dots$ and the normal curve generating the expansion is allowed to have the same mean and variance as z_N , we have that the standardized deviation from the unit normal deviate ξ , say, of the percentage point, is given by

$$\frac{z_N - \kappa_{1N}}{(\kappa_{2N})^{\frac{1}{2}}} - \xi = \frac{1}{6} l_3 (\xi^2 - 1) + \frac{1}{24} l_4 (\xi^3 - 3\xi) - \frac{1}{36} l_3^2 (2\xi^3 - 5\xi) + \dots,$$

where

$$l_r = \frac{\kappa_{rN}}{\kappa_{2N}^{\frac{1}{2}r}} \quad (r > 2).$$

Provided the moments of z are not greatly affected by non-normality or by unequal group frequencies there is no reason why a Fisher-Cornish expansion should not be adequate in the non-normal case. We shall assume therefore that the departures are of such order for this to be true. If $L_1, L_2, \dots$ refer to non-normal z, z_A (say), then let

$$L_1 = \frac{\kappa_{1A} - \kappa_{1N}}{(\kappa_{2N})^{\frac{1}{2}}}, \quad L_2 = \frac{\kappa_{2A} - \kappa_{2N}}{\kappa_{2N}}, \quad L_r = \frac{\kappa_{rA}}{(\kappa_{2N})^{\frac{1}{2}r}} \quad (r > 2).$$

In this case the Fisher-Cornish expansion is

$$\frac{z_A - \kappa_{1N}}{\sqrt{\kappa_{2N}}} - \xi = L_1 + \frac{1}{6}L_3(\xi^2 - 1) + \frac{1}{2}L_2\xi + \frac{1}{24}L_4(\xi^3 - 3\xi) - \frac{1}{36}L_3^2(2\xi^3 - 5\xi) + \dots$$

Hence by subtraction

$$\begin{aligned} \frac{z_A - z_N}{\sqrt{\kappa_{2N}}} = \frac{\kappa_{1A} - \kappa_{1N}}{\sqrt{\kappa_{2N}}} + \frac{\xi}{2} \frac{(\kappa_{2A} - \kappa_{2N})}{\kappa_{2N}} + \frac{(\xi^2 - 1)}{6} \left(\frac{\kappa_{3A} - \kappa_{3N}}{\kappa_{2N}^{\frac{3}{2}}} \right) \\ + \frac{(\xi^3 - 3\xi)}{24} \left(\frac{\kappa_{4A} - \kappa_{4N}}{\kappa_{2N}^2} \right) - \frac{1}{36} (2\xi^3 - 5\xi) \left[\frac{\kappa_{3A}^2 - \kappa_{3N}^2}{\kappa_{2N}^3} \right] + \dots \end{aligned} \quad (1)$$

If desired the cumulants of z_N can be found from the differentiation of the cumulant generating function (see, for example, Wishart (1947))

$$\begin{aligned} \mathcal{K} = \log \left(\frac{f_1 + it - 2}{2} \right)! - \log \left(\frac{f_1 - 2}{2} \right)! - \frac{1}{2}it \log \left(\frac{1}{2}f_1 \right) \\ - \log \left(\frac{f_2 + it - 2}{2} \right)! + \log \left(\frac{f_2 - 2}{2} \right)! + \frac{1}{2}it \log \left(\frac{1}{2}f_2 \right), \end{aligned}$$

for the polygamma functions are well tabulated, but this is not really necessary for the difference. We also note that z_N is the tabulated percentage point of z , given, for instance, in Fisher and Yates' Tables. z_A can therefore be obtained provided the differences of the cumulants of z_A and z_N are known to a sufficient order of approximation. κ_{2N} is of order n^{-1} and l_r and L_r are of order $n^{1-\frac{1}{2}r}$ when the group frequencies are equal.

By definition

$$2z = \log F = \log (S_1/f_1) - \log (S_2/f_2)$$

or

$$2z = \log \left(1 + \frac{S_1 - \mathcal{E}(S_1)}{\mathcal{E}(S_1)} \right) - \log \left(1 + \frac{S_2 - \mathcal{E}(S_2)}{\mathcal{E}(S_2)} \right).$$

Write δ_1 and δ_2 for reduced, standardized, S_1 and S_2 , respectively, and we have

$$2z = (\delta_1 - \delta_2) - \frac{1}{2}(\delta_1^2 - \delta_2^2) + \frac{1}{3}(\delta_1^3 - \delta_2^3) - \frac{1}{4}(\delta_1^4 - \delta_2^4) + \dots$$

δ_1 and δ_2 are the two quantities the cumulants of which were given earlier. Taking expectations and remembering that those terms remaining when the γ 's are put equal to zero are those belonging to twice the mean value of normal z we have

$$\kappa_{1A} - \kappa_{1N} = -\frac{\gamma_2}{4n}(V_{012} - V_{210}) + \left\{ \frac{\gamma_4}{6n^2}(V_{023} - V_{320}) - \frac{3}{8} \frac{\gamma_2^2}{n^2}(V_{012}^2 - V_{210}^2) + \frac{\gamma_2}{2n} \left(\frac{V_{012}}{f_1} - \frac{V_{210}}{f_2} \right) \right\} + \dots$$

to order n^{-2} . Following the usual procedure we obtain for the higher cumulants,

$$\begin{aligned}\kappa_{2A} - \kappa_{2N} &= \frac{\gamma_2}{4n} (V_{012} - 2V_{111} + V_{210}) + \left\{ -\frac{\gamma_4}{4n^2} (V_{023} - V_{122} - V_{221} + V_{320}) \right. \\ &\quad + \frac{\gamma_2^2}{4n^2} \left(\frac{5}{2} V_{012}^2 + \frac{5}{2} V_{210}^2 - 2V_{111} V_{012} - 2V_{111} V_{210} - V_{111}^2 \right) \\ &\quad \left. - \frac{\gamma_2}{8n} \left(V_{012} \left(\frac{3}{f_1} + \frac{1}{f_2} \right) + V_{210} \left(\frac{1}{f_1} + \frac{3}{f_2} \right) \right) \right\} + \dots, \\ \kappa_{3A} - \kappa_{3N} &= \frac{\gamma_4}{8n^2} (V_{023} - 3V_{122} + 3V_{221} - V_{320}) - \frac{3}{8} \frac{\gamma_2^2}{n^2} (V_{012} - V_{210}) (V_{012} + V_{210} - 2V_{111}) + \dots, \\ \kappa_{4A} - \kappa_{4N} &= \frac{\gamma_6}{16n^3} (V_{034} - 4V_{133} + 6V_{232} - 4V_{331} + V_{430}) \\ &\quad + \frac{\gamma_4 \gamma_2}{8n^3} (V_{023} (-5V_{012} + 6V_{111} - V_{210}) + V_{320} (-5V_{210} + 6V_{111} - V_{120}) \\ &\quad + V_{122} (9V_{012} - 3V_{210} - 6V_{111}) + V_{221} (9V_{210} - 3V_{012} - 6V_{111})) \\ &\quad + \frac{\gamma_2^3}{4n^3} (5V_{012}^3 + 5V_{210}^3 + 2V_{111}^3 + 3V_{111}^2 (V_{012} + V_{210}) \\ &\quad - 12V_{111} (V_{012}^2 + V_{210}^2) + 6V_{111} V_{210} V_{012}) \\ &\quad + \frac{\gamma_4}{4n^2} \left(V_{122} \left(\frac{3}{f_1} - \frac{6}{f_2} \right) + V_{221} \left(\frac{3}{f_2} - \frac{6}{f_1} \right) + V_{111} \frac{6n}{f_1 f_2} \right) \\ &\quad + \frac{\gamma_2^2}{2n^2} \left(-\frac{3}{f_1} (V_{012} - V_{111})^2 - \frac{3}{f_2} (V_{210} - V_{111})^2 + \frac{1}{f_1} \left(4V_{023} - 6V_{122} + 9V_{112} \frac{n}{f_2} \right) \right. \\ &\quad \left. + \frac{1}{f_2} \left(6V_{320} - 14V_{221} + 3V_{211} \frac{n}{f_1} \right) + \frac{n}{f_2^2} \left(4V_{111} - 3V_{210} + \frac{n}{f_2} \right) \right).\end{aligned}$$

The second and third cumulants are of order n^{-2} and the fourth is of order n^{-3} .

If the numbers in the groups are all equal, the V 's are all unity and considerable simplification results. We have

$$\begin{aligned}\kappa_{1A} - \kappa_{1N} &= \frac{\gamma_2}{2n} \left(\frac{1}{f_1} - \frac{1}{f_2} \right) + \dots \\ \kappa_{2A} - \kappa_{2N} &= -\frac{\gamma_2}{2n} \left(\frac{1}{f_1} + \frac{1}{f_2} \right) + \dots \\ \kappa_{3A} - \kappa_{3N} &= 0 + \dots \\ \kappa_{4A} - \kappa_{4N} &= \frac{\gamma_4}{4n^2} \left(-\frac{3}{f_1} - \frac{3}{f_2} + \frac{6n}{f_1 f_2} \right) + \frac{\gamma_2^2}{2n^2} \left(-\frac{2}{f_1} - \frac{8}{f_2} + \frac{12n}{f_1 f_2} + \frac{n}{f_2^2} + \frac{n^2}{f_2^3} \right) + \dots\end{aligned}$$

It is possible to rewrite the moments in a slightly different way which will bring out more strongly the variability of the group frequencies and the effect which it has on the moments of z . Let, for $i = 1, 2, \dots, k$,

$$n_i = \frac{n}{k} + \Delta_i;$$

then

$$\sum_{i=1}^k \Delta_i^2 = k \text{var}(n_i) = kv \quad (\text{say}).$$

We take the expression for V_{abc} , substitute for n_i , and expand to get

$$V_{abc} - 1 = v \left[{}^a C_2 \frac{k^2}{f_2^2} - \frac{ack^2}{nf_1 f_2} + {}^c C_2 \frac{k^2}{n^2 f_1^2} + {}^{b+1} C_2 \frac{k^2}{n^2} - \frac{abk^2}{nf_2} + \frac{bck^2}{n^2 f_1} \right] \\ + \mu_3(n) \left[{}^a C_3 \frac{k^3}{f_2^3} - {}^a C_2 {}^c C_1 \frac{k^3}{nf_1 f_2^2} + {}^c C_2 {}^a C_1 \frac{k^3}{n^2 f_1^2 f_2} - {}^c C_3 \frac{k^3}{n^3 f_1^3} - {}^{b+2} C_3 \frac{k^3}{n^3} + {}^{b+1} C_2 {}^a C_1 \frac{k^3}{n^2 f_2} \right. \\ \left. - {}^{b+1} C_2 {}^c C_1 \frac{k^3}{n^2 f_1} - {}^a C_2 {}^b C_1 \frac{k^3}{nf_2^2} + {}^a C_1 {}^b C_1 {}^c C_1 \frac{k^3}{n^2 f_1 f_2} - {}^b C_1 {}^c C_2 \frac{k^3}{n^3 f_1^2} \right] + \dots$$

$$V_{111} - 1 = -\frac{vk^4}{n^2 f_1 f_2} + \frac{k^5}{n^3 f_1 f_2} \mu_3(n).$$

For example

If $k = 5$ and $n = 20$ this gives us

$$V_{111} - 1 = -\frac{5}{192}v + \frac{5}{768}\mu_3(n)$$

which suggests that for all but widely different group frequencies the first term involving v will be adequate, and this is so in fact. We therefore substitute for V_{abc} and collect terms which enables the effect of the unequal group frequencies, as described by their variance, to be seen immediately

Let

$$A_r = k^r \left(\frac{1}{f_1^r} - \frac{1}{f_2^r} \right)$$

$$B_r = k^r \left(\frac{1}{f_1^r} + \frac{1}{f_2^r} \right).$$

Then on substitution for V_{abc} we have

$$\kappa_{1A} - \kappa_{1N} = -\frac{\gamma_2 k^2 A_2 v}{4n^3} + \left\{ \frac{\gamma_2 A_1}{2nk} + \frac{v k^2}{2n^4} \left(\gamma_4 A_2 - \frac{3}{2} \gamma_2^2 A_2 + \gamma_2 \frac{n}{k} A_3 \right) - \frac{3}{8} v^2 \frac{k^4}{n^6} \gamma_2^2 A_4 \right\} + \dots$$

$$\kappa_{2A} - \kappa_{2N} = \frac{\gamma_2 k^2 B_1^2 v}{4n^3} + \left\{ -\frac{\gamma_2 B_1}{2nk} + \frac{vk^2 B_1^2}{4n^4} (-2\gamma_4 + 3\gamma_2^2) + \frac{vk}{4n^3} \gamma_2 (B_1 B_2 - A_1 A_2) \right. \\ \left. + \frac{v^2 k^4 \gamma_2^2}{4n^6} \left(\frac{1}{4} B_1^2 - \frac{9k^2}{f_1 f_2} B_1^2 - \frac{A_2^2}{4} \right) \right\} + \dots$$

$$\kappa_{3A} - \kappa_{3N} = -\frac{3}{8} v^2 \frac{k^4}{n^6} B_1^2 A_2 \gamma_2^2 + \dots$$

$$\kappa_{4A} - \kappa_{4N} = -\frac{3}{4} \frac{\gamma_4 \gamma_2}{n^3} \frac{k^4}{n^4} B_1^4 v^2 \\ + \frac{\gamma_2^3}{4n^3} \left(\frac{3vk^4}{n^4} \left(B_2^2 + \frac{8k^2}{f_1 f_2} B_2 \right) + \frac{v^3 k^6}{n^6} \left(5B_6 - \frac{8k^6}{f_1^3 f_2^3} + \frac{3k^4}{f_1^2 f_2^2} B_2 + \frac{12k^2}{f_1 f_2} B_4 \right) \right) \\ + \frac{\gamma_4}{4n^2} \left(-3 \left(\frac{1}{f_1} + \frac{1}{f_2} \right) + \frac{6n}{f_1 f_2} + \frac{kv}{n^2} \left(3B_3 - \frac{6nk^3}{f_1^2 f_2^2} \right) \right) \\ + \frac{\gamma_2^2}{2n^2} \left(-\frac{2}{f_1} - \frac{8}{f_2} + \frac{12n}{f_1 f_2} + \frac{n}{f_2^2} + \frac{n^2}{f_2^3} + v \left(\frac{4k}{n^2} B_1^3 + \frac{k^4}{n^2} \left(\frac{2}{f_1^3} + \frac{16}{f_1 f_2^2} \right) \right) \right. \\ \left. + \frac{k^2}{f_1 f_2} \left(-\frac{12}{f_1 f_2} - \frac{12k}{f_1 f_2} + \frac{k^2}{n} \left(\frac{9}{f_1^2} + \frac{3}{f_2^2} \right) \right) + v^2 \left(-\frac{3k^3}{n^4} B_1^2 B_3 \right) \right).$$

When $v = 0$, that is when all the numbers in the groups are equal, the cumulants reduce to those previously noted.

We consider the case of equal group frequencies first. Refer to equation (1), substitute in the cumulant differences and collect terms of like order. The difference between the tabulated value of *z* and that of non-normal *z* is to order $n^{-\frac{3}{2}}$ and for *k* reasonably large

$$-\frac{1}{\sqrt{2}} \frac{\gamma_2}{n} \left(\frac{B_1}{k}\right)^{\frac{1}{2}} \frac{\xi}{2} + \left(\frac{k}{B_1}\right)^{\frac{3}{2}} \frac{\sqrt{2}}{n^2} \frac{(\xi^3 - 3\xi)}{24} \left[\gamma_2^2 \left(-\frac{2}{f_1} - \frac{8}{f_2} + \frac{12n}{f_1 f_2} + \frac{n}{f_2^2} + \frac{n^2}{f_2^3} \right) + \frac{\gamma_4}{2} \left(-\frac{3}{f_1} - \frac{3}{f_2} + \frac{6n}{f_1 f_2} \right) \right]. \quad (2)$$

On the whole this correction is not large. For $n = 20$, $k = 5$ it is

$$-\gamma_2 \frac{\xi}{2} 0.020 + \frac{1}{24} (\xi^3 - 3\xi) (\gamma_2^2 0.063 + \gamma_4 0.010)$$

while for $n = 20$ and $k = 10$ it is

$$-\gamma_2 \frac{\xi}{2} 0.015 + \frac{1}{24} (\xi^3 - 3\xi) (\gamma_2^2 0.086 + \gamma_4 0.013).$$

For the sake of illustration consider the rectangle, where $\gamma_2 = -1.2$, $\gamma_4 = 48/7$. The alterations to the tabulated percentage points are as follows

Degrees of freedom	Normal $z_{0.05}$	Cor-rection	Normal $z_{0.01}$	Cor-rection
$f_1 = 4, f_2 = 15$	0.559	0.016	0.794	0.066
$f_1 = 9, f_2 = 10$	0.553	0.011	0.799	0.072

The corrections are not, on the whole, large and are unlikely to cause serious error in the application of tests of significance. The stability of the 5% significance level is worthy of note, since it has become apparent in many non-normal investigations. (See, for example, Pearson and Merrington's table 42 in *Biometrika Tables for Statisticians*.)

It will be noticed that while the effect of non-normality with equal groups is to alter the mean and variance of normal *z* by terms of order n^{-2} the effect of unequal group frequencies is to alter them by terms of order n^{-1} . This means that it is no longer possible to expand about moments of z_N since the Edgeworth expansion requires the assumption that the difference between the second cumulants of z_A and z_N is of order n^{-2} . The simplest thing would seem to be therefore to expand the mean and variance of z_A with unequal group frequencies (z_A^* say) about its own mean and variance, remembering that the cumulants of z_N which can be substituted are given by

$$2\kappa_{1N} = \log f_2 - \log f_1 + \psi(\tfrac{1}{2}f_1) - \psi(\tfrac{1}{2}f_2)$$

$$2^s \kappa_{sN} = \psi^{s-1}(\tfrac{1}{2}f_1) + (-1)^s \psi^{s-1}(\tfrac{1}{2}f_2) \quad (s > 1),$$

where ψ is the digamma function and ψ^{s-1} is its $(s-1)$ st differential. For moderate divergencies from the ideal equal group frequency situation, such as envisaged in the expansions for the moments in terms of *v* given earlier, it is clear from numerical calculation that the effect of these divergencies is to bring the mean and variance of z_A^* back towards those of z_N . The greatest contributions to the mean and to the variance are, in fact,

$$\kappa_{1A^*} = \kappa_{1N} - \gamma_2 \left(\frac{k^2 A_2 v}{4n^3} - \frac{A_1}{2nk} \right),$$

$$\kappa_{2A^*} = \kappa_{2N} + \gamma_2 \left(\frac{k^2 B_1^2 v}{4n^3} - \frac{B_1}{2nk} \right).$$

It is suggested therefore that calculations for the case when the group frequencies are equal, i.e. when $v = 0$, will give a reasonable idea of the magnitude of the error involved if there is non-normality.

The purpose of this present note has been one of method, but E. S. Pearson has pointed out to me that Gayen (1950) also considered the F -test in the one-way classification in some detail. Gayen was concerned in his numerical examples with the effect on the nominal 5% F -level of non-normality and unequal group frequencies and his results are not therefore directly comparable with those obtained here. It is, however, possible to use the same technique to obtain corrections to the tail areas as has been used above to obtain corrections to the percentage point. Thus if $g(\eta)$ is the normal ordinate at η , then for the equal groups case the correction to

$$\int_{-\infty}^{\eta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$$

will be found to be, to order n^{-1} ,

$$g(\eta) \left[-\frac{\eta}{2} \frac{(\kappa_{2A} - \kappa_{2N})}{\kappa_{2N}} + \frac{(\eta^3 - 3\eta)}{24} \frac{(\kappa_{4A} - \kappa_{4N})}{\kappa_{2N}^2} \right]. \quad (3)$$

For reasonable sized k we may write this as

$$g(\eta) \left[\frac{\eta}{2} \frac{\gamma_2}{n} + \frac{(\eta^3 - 3\eta)}{24} \frac{k^2}{B_1^2} \left(\frac{\gamma_4}{n^2} \left(-\frac{3}{f_1} - \frac{3}{f_2} + \frac{6n}{f_1 f_2} \right) + \frac{2\gamma_2^2}{n^2} \left(-\frac{2}{f_1} - \frac{8}{f_2} + \frac{12n}{f_1 f_2} + \frac{n}{f_2^2} + \frac{n^2}{f_2^3} \right) \right) \right].$$

Somewhat surprisingly, the major part of the correction does not appear to depend on the degrees of freedom, although these are implicit in η ; for η is equal to the standardized percentage point of z_N , i.e.

$$\eta = \frac{z_N(\alpha) - \kappa_{1N}}{\sqrt{\kappa_{2N}}}.$$

The comparison with Gayen's work is difficult in that he assumes for the parent population itself an Edgeworth series in which the terms involving γ_4 (of the population) and γ_2^2 are assumed not to exist. It would appear therefore that a comparison should only be made in the region in which the parent population of Gayen is positive definite, i.e. is capable of being a probability distribution. I have therefore chosen a Pearson Type VII for my population with $\gamma_2 = 1$ and $\gamma_4 = 10$. [This is a t -distribution with 10 d.f.] Let the degrees of freedom of the samples be $f_1 = 8, f_2 = 12$. Gayen's table will give the correction to be made to the 0.05 upper level as -0.0030 . Substitution in formula (3) gives the correction as -0.002 (3). The corrections are therefore of the same order, and (3) can be used to obtain some idea of the error involved for any of the usual significance levels for reasonable k .

I have to thank E. S. Pearson for pointing out errors in the original draft, and H. Scheffé whose remarks on the paucity of arithmetic spurred me into trying to remedy the situation.

REFERENCES

- DAVID, F. N. & JOHNSON, N. L. (1951a). *Biometrika*, **38**, 43-57.
 DAVID, F. N. & JOHNSON, N. L. (1951b). *Ann. Math. Statist.* **22**, 382-93.
 DAVID, F. N. & JOHNSON, N. L. (1952). *Ann. Math. Statist.* **23**, 594-602.
 GAYEN, A. K. (1950). *Biometrika*, **37**, 236-55.
 PEARSON, E. S. (1931). *Biometrika*, **23**, 114-33.
 WISHART, J. (1947). *Biometrika*, **34**, 170-8.

APPROXIMATION TO THE DISTRIBUTION OF SAMPLE SIZE FOR SEQUENTIAL TESTS

I. TESTS OF SIMPLE HYPOTHESES

By D. H. BHATE

University College London and Delhi School of Economics, New Delhi

1. INTRODUCTION

A. Wald (1943) developed a sequential probability ratio test for testing a simple hypothesis against a simple alternative. He gave approximate formulae for the sequential test procedure, its average sample size, and the power of the test. He has also derived the asymptotic form of the distribution of the sample size. The actual distribution of the decisive sample number (D.S.N.) has proved difficult to determine. Kac (1945) gave the distribution of the D.S.N. for the test approximately equivalent to that for comparing two possible values for the mean of a Normal distribution, in terms of the latent roots of an integral equation. Darling & Siegert (1953) have given the distribution of the duration of an absorption process approximating the sequential procedure for testing the mean of a Normal population with known variance. But in practice it is rather difficult to evaluate the actual probability of reaching a decision at a specified value of n . Baker (1950) has studied the distribution of D.S.N. experimentally in a special case. Page (1954*a*) and Albert (1954) have also studied this problem, using Monte Carlo methods and similar techniques. Ramchandran (1954) has developed certain approximations to the distribution of the D.S.N. His method gives satisfactory results for small values, only, of n , because the errors of approximation accumulate with n .

In this paper we shall derive general formulae for upper and lower bounds for the cumulative distribution of the D.S.N., and use them to estimate its approximate distribution.

2. DEVELOPMENT OF THEORY

2.1. Let $f(x, \theta)$ be the probability density function of a random variable x , depending on a parameter θ . A sequential procedure, for testing the hypothesis $H_0: \theta = \theta_0$ against the alternative $H_1: \theta = \theta_1$ (H_0 and H_1 may be either simple or composite) may be constructed as follows.

Suppose that $x_1, x_2, \dots, x_j$ is a sequence of independent observations and that t_j , a function of these x 's, is an appropriate test statistic. Then the sequential test procedure is based on a partition of the range of variation of t_j into three mutually exclusive regions C_j , A_j and R_j , using the following rule:

- (i) if $t_j \in C_j$, take a further observation;
- (ii) if $t_j \in A_j$, accept H_0 ;
- (iii) if $t_j \in R_j$, reject H_0 (i.e. accept H_1).

The regions are chosen in such a way that the probabilities of rejecting H_0 and H_1 when they are true are approximately α and β , respectively.

Following this rule a decision is reached at the i th stage if $t_j \in C_j$ for $j = 1, 2, \dots, i-1$, and either $t_i \in A_i$ or $t_i \in R_i$. The argument used below to determine the upper and lower bounds

for the distribution of i , the decisive sample number (D.S.N.) may be illustrated with the help of the accompanying diagram, in which j is taken as abscissa and t_j as the ordinate. The boundaries between the regions have been represented by the lines $O'A$, and $O''R$ (not necessarily straight lines).

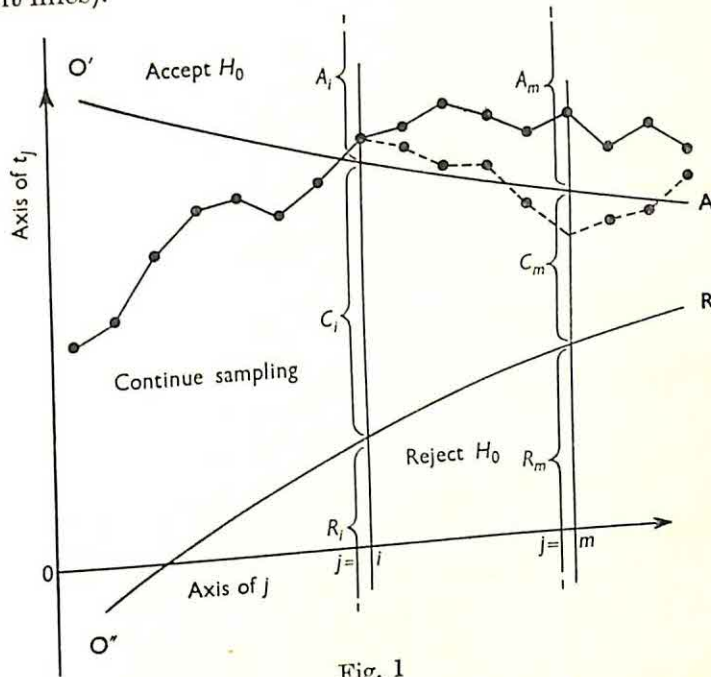


Fig. 1

2.2. Suppose now that although a decision is reached at stage $j = i$, sampling is continued to a fixed stage $j = m$ ($\geq i$). In the diagram, two possible sequences of samples are represented. Each track enters the acceptance region at $j = i$ but in one case $t_m \in A_m$ and in the other case $t_m \in C_m$. Similar sequences of samples when the track enters the rejection region at $j = i$ are also possible.

Now define the following probabilities.

- (a) P_m = probability that a decision is reached at, or before the m th state and $t_m \in C_m = P\{t_m \in C_m\}$.
- (b) p_i = probability that a decision is reached at the i th stage (and not before) and $t_i \in R_i\} = p_{iA} + p_{iR}$ (say).
- (c) p'_i = probability that a decision is reached at the i th stage (and not before), but that when sampling is continued to the m th stage, $t_m \in C_m$
- $$= P\{t_j \in C_j \text{ for } j = 1, 2, \dots, i-1, t_i \in A_i \text{ and } t_m \in C_m\}$$
- $$+ P\{t_j \in C_j \text{ for } j = 1, 2, \dots, i-1, t_i \in R_i \text{ and } t_m \in C_m\}$$
- $$= p'_{iA} + p'_{iR} \text{ (say).}$$

The diagram may help to clarify the meaning of these probabilities. The continuous path would be included among those associated with p_{iA} and P_m but not with p'_{iA} ; the broken path would be included among those associated with p_{iA} and p'_{iA} but not with P_m .

Then it follows that

$$\sum_{i=1}^m p_i = P_m + \sum_{i=1}^{m-1} (p'_{iA} + p'_{iR}).$$

(2.2.1)

Thus P_m provides a lower limit for the cumulative distribution of the decisive sample number. By considering the upper bound for p'_{iA} and p'_{iR} we shall obtain an upper limit for the cumulative distribution of the D.S.N.

Let

$$p_i(t_i) = p(t_i | t_j \in C_j \text{ for } j = 1, 2, \dots, i-1)$$

be the conditional probability density function of t_i , given that no decision has been reached before the i th stage.

Then

$$p_{iA} = \int_{t_i \in A_i} p_i(t_i) dt_i \quad \text{and} \quad p_{iR} = \int_{t_i \in R_i} p_i(t_i) dt_i.$$

Further

$$p'_{iA} = \int_{t_i \in A_i} p_i(t_i) \left\{ \int_{t_m \in C_m} p(t_m | t_i) dt_m \right\} dt_i,$$

where $p(t_m | t_i)$ is the conditional distribution of t_m given t_i . If we write

$$\eta_{im} = \max_{t_i \in A_i} \int_{t_m \in C_m} p(t_m | t_i) dt_i, \quad (2.2.2)$$

it follows that

$$p'_{iA} \leq \eta_{im} \int_{t_i \in A_i} p_i(t_i) dt_i = \eta_{im} p_{iA}.$$

Similarly if

$$\delta_{im} = \max_{t_i \in R_i} \int_{t_m \in C_m} p(t_m | t_i) dt_i. \quad (2.2.3)$$

$$p'_{iR} \leq \delta_{im} p_{iR}.$$

Hence from (2.2.1) we have the bounds

$$P_m \leq \sum_{i=1}^m p_i \leq P_m + \sum_{i=1}^{m-1} \eta_{im} p_{iA} + \sum_{i=1}^{m-1} \delta_{im} p_{iR} \quad (2.2.4)$$

The result is valid for any hypothesis H (not only for H_0 and H_1).

Two less sharp, but sometimes more easily calculable, inequalities can be derived from (2.2.4). These are

$$\sum_{i=1}^m p_i \leq P_m + \sum_{i=1}^{m-1} p_i \eta_{im} + \alpha \max_{i,m} \delta_{im} \quad \text{when } H_0 \text{ is true,} \quad (2.2.5)$$

$$\sum_{i=1}^m p_i \leq P_m + \beta \max_{i,m} \eta_{im} + \sum_{i=1}^{m-1} p_i \delta_{im} \quad \text{when } H_1 \text{ is true} \quad (2.2.6)$$

We can use (2.2.4) to give useful upper and lower bounds for the cumulative distribution of D.S.N. if η_{im} and δ_{im} can be evaluated, and, being evaluated, are sufficiently small.

Just where, between these limits, the actual value of the cumulative distribution lies, we do not know, but in the examples which follow it appears that a simple average of the upper and the lower bounds gives quite good results.

2.3. Consider now a special case of comparison of two simple hypotheses H_0 ; H_1 and suppose that we know that, with an obvious notation

$$p_i(H_0) = p_i(H_1).$$

This is so, for example, when H_0 and H_1 specify two values for the mean of a Normal population, and $\alpha = \beta$.

For all decisions to accept H_0 , at sample size i

$$p(x_1, \dots, x_i | H_1) < \frac{\beta}{1-\alpha} p(x_1, \dots, x_i | H_0),$$

so

$$p_{iA}(H_1) < \frac{\beta}{1-\alpha} p_{iA}(H_0)$$

and similarly

$$p_{iR}(H_1) > \frac{1-\beta}{\alpha} p_{iR}(H_0).$$

Approximately (i.e. neglecting the amount by which the limits are overshoot) the inequalities may be replaced by equalities, giving, when $\alpha = \beta$,

$$\frac{p_{iA}(H_1)}{p_{iA}(H_0)} = \frac{p_{iR}(H_0)}{p_{iR}(H_1)} = \frac{\alpha}{1-\alpha}.$$

Since

$$\begin{aligned} p_{iA}(H_1) + p_{iR}(H_1) &= p_i(H_1) = p_i(H_0) = p_{iA}(H_0) + p_{iR}(H_0), \\ p_{iA}(H_1) &\div \alpha p_i \quad \text{and} \quad p_{iA}(H_0) = (1-\alpha) p_i, \\ p_{iR}(H_1) &\div (1-\alpha) p_i \quad \text{and} \quad p_{iR}(H_0) = \alpha p_i. \end{aligned} \quad (2.3.1)$$

If (2.3.1) be assumed to be true we can proceed to use the inequalities (2.2.4).

Table 1

i	c	i	c	i	c
1	0.0860	6	0.0687	11	0.0686
2	.0778	7	.0686	12	.0686
3	.0710	8	.0686	13	.0686
4	.0693	9	.0686	14	.0686
5	.0688	10	.0686	15	.0686

In order to see how far this assumption is justified we quote figures obtained by quadrature in the course of the calculation of the exact distribution of the D.S.N. for the case

$$\left\{ \begin{aligned} p(x | \theta) &= \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}(x-\theta)^2} \\ H_0: \theta &= \theta_0 = 0, \quad H_1: \theta = \theta_1 = 1; \quad \alpha = \beta = 0.0758. \end{aligned} \right.$$

(The common value 0.0758 for α and β was chosen in Example 2 below, to give simple limits for $t_m = \sum_{j=1}^m x_j$.)

The values of the ratio $c = p_{iR}(H_0)/p_i(H_0)$ obtained in this calculation are given in table 1. Apart from some variation for small i , when some excess of the observed statistic over the boundaries is to be expected, $c = p_{iR}(H_0)/p_i(H_0) \div 0.0686$. Thus the assumption of constancy in the ratio may not be greatly in error, provided the probability of termination for $i = 1, 2, 3, 4$, is fairly small (though the actual value of the ratio is rather less than the nominal value of α).

3. ILLUSTRATIVE EXAMPLES

Example 1. Let

$$f(x, \theta) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}(x-\theta)^2},$$

$$H_0: \theta = \theta_0 = 0, \quad H_1: \theta = \theta_1 = 1; \quad \alpha = \beta = 0.025.$$

(i) The continuation region at the m th stage, C_m , is defined by the inequalities

$$\frac{0.025}{1-0.025} < \prod_{i=1}^m \frac{f(x_i, 1)}{f(x_i, 0)} < \frac{1-0.025}{0.025},$$

or

$$-\log_e 39 < -\frac{1}{2} \sum_{i=1}^m [(x_i - 1)^2 - x_i^2] < \log_e 39$$

i.e.

$$-3.66 + \frac{1}{2}m < t_m < 3.66 + \frac{1}{2}m,$$

where

$$t_m = \sum_{i=1}^m x_i.$$

Similarly

$$(ii) \quad -3.66 + \frac{1}{2}m \geq t_m \quad \text{for region } A_m,$$

$$(iii) \quad 3.66 + \frac{1}{2}m \leq t_m \quad \text{for region } R_m.$$

Denoting the probability density function of t_m by $p(t_m)$, where in this case

$$p(t_m) = \frac{1}{\sqrt{(2\pi m)}} \exp \left\{ -\frac{1}{2}(t_m - m\theta)^2/m \right\},$$

we have

$$\eta_{im} = \int_{\frac{1}{2}(m-i)}^{7.32 + \frac{1}{2}(m-i)} p(t_{m-i}) dt_{m-i},$$

since this value corresponds to the case when t_i is exactly on the boundary of C_i . Similarly

$$\delta_{im} = \int_{-7.32 + \frac{1}{2}(m-i)}^{\frac{1}{2}(m-i)} p(t_{m-i}) dt_{m-i}.$$

The following Table 2 gives, when H_0 is true, the upper and lower bounds for the cumulative distribution of the D.S.N., using (2.2.4), and the average of these bounds. The observed frequencies in the last column are taken from an experimental investigation by Baker (1950).

Comparing the theoretical frequencies Np_i with Baker's sampling results, we find that $\chi^2 = 19.0022$ with $\nu = 20$ D.F. (frequencies corresponding to $i = 1$ and 2 were grouped). Thus the somewhat arbitrary assumption that we may approximate to the cumulative sum of the p_i 's by taking the mid-point between the upper and lower bounds is not disproved by these results.

Example 2

Let $\alpha = \beta = 0.0758$. This case has already been referred to in § 2. The regions C_m, A_m, R_m are now defined by

$$C_m: -2.50 + \frac{1}{2}m < t_m < 2.50 + \frac{1}{2}m,$$

$$A_m: -2.50 + \frac{1}{2}m \geq t_m,$$

$$R_m: 2.50 + \frac{1}{2}m \leq t_m.$$

The approximate distribution of the D.S.N. under H_0 can be calculated as in Example 1. In this case, an exact calculation by quadrature was also carried out. At each stage $p_c(t_m)$, the conditional probability density function of t_m , given that no decision had been

reached up to and including the m th stage, was computed by quadrature. The distribution of $t_{m+1} = t_m + x_{m+1}$ was then computed from the formula

$$p_c(t_{m+1}) = \int_{t_m \in C_m} p_c(t_m) p(x_{m+1}) dt_m, \quad \text{where } x_{m+1} = t_{m+1} - t_m.$$

The results of these calculations are shown in Table 3, together with some observed frequencies taken from the unpublished results of experiments carried out by Baker, who kindly put them at my disposal.

Table 2. Estimation of p_i for Example 1 and comparison with Baker's empirical results with $N = 2002$

Sample no. i	Lower bound L_i	Upper bound U_i	$\frac{1}{2}(U_i + L_i)$	Diff. = $\hat{p}_i$	Expected frequency $N\hat{p}_i$	Observed frequency
				0.00080	2	0
1	0.00080	0.00080	0.00080	.02988	60	60
2	.03055	.03080	.03068	.08126	163	140
3	.10709	.11680	.11194	.11043	221	217
4	.20560	.23914	.22237	.11311	226	213
5	.30451	.36645	.33548			
				0.10437	209	219
6	0.39684	0.48287	0.43985	.09131	183	193
7	.47944	.58289	.53116	.07709	154	158
8	.55112	.66538	.60825	.06495	130	125
9	.61352	.73289	.67320	.05285	106	112
10	.66583	.78627	.72605			
				0.04469	89	99
11	0.71173	0.82976	0.77074	.04064	81	85
12	.75438	.86839	.81138	.03107	62	63
13	.78764	.89726	.84245	.02455	49	54
14	.81545	.91855	.86700	.02178	44	43
15	.84090	.93667	.88878			
				0.01740	35	41
16	0.86174	0.95062	0.90618	.01527	31	21
17	.88059	.96231	.92145	.01367	27	37
18	.89761	.97263	.93512	.01045	21	26
19	.91114	.98000	.94557	.00906	18	15
20	.92334	.98592	.95463			
				0.02807	56	54
21-25	0.93419	—	—	.01730	35	27
> 25	.96838	—	—			

Comparing the approximate expected frequencies $N\hat{p}_i$ and the exact expected frequencies Np_i with the observed frequencies we find that $\chi^2 = 18.7304$ and $\chi^2 = 13.4573$, respectively, for $\nu = 16$. Thus we find that there is agreement between the expected and the observed frequencies. It is clear from this table that the approximate expected frequencies and the exact values of these frequencies agree well. This supports the assumption that we may approximate the cumulative distribution by the mean of the lower and upper bounds.

Table 3. Values p_i obtained by (2.2.4) and by quadrature and their comparison with empirical results with $N = 698$

Sample no. i	Approximate probabilities $\hat{p}_i$	Exact probabilities p_i	Expected frequencies		Observed frequency
			$N\hat{p}_i$	Np_i	
1	0.02410	0.02410	17	17	17
2	.13126	.13355	92	93	100
3	.16425	.16311	115	114	111
4	.14608	.14459	102	101	100
5	.11702	.11735	82	82	94
6	0.09112	0.09257	64	65	66
7	.07080	.07229	49	50	45
8	.05613	.05628	39	39	37
9	.04326	.04376	30	31	41
10	.03489	.03399	24	24	18
11	0.02796	0.02643	19	18	13
12	.02258	.02052	16	14	12
13	.01844	.01594	13	11	13
14	.01526	.01238	11	9	8
15	.01236	.00963	9	7	7
16	0.01026	0.00747	7	5	3
≥ 17	.01423	.02604	10	18	13

Example 3

Let x be a Normal variate with zero mean and variance σ^2 .

$$H_0: \sigma = \sigma_0 = 1, \quad H_1: \sigma = \sigma_1 = 3; \quad \alpha = \beta = 0.01.$$

The test procedure is then based on the following regions:

$$C_m: -10.350 + 2.475m < \sum_{j=1}^m x_j^2 < 10.350 + 2.475m,$$

$$A_m: -10.350 + 2.475m \geq \sum_{j=1}^m x_j^2,$$

$$R_m: 10.350 + 2.475m \leq \sum_{j=1}^m x_j^2.$$

In this case we cannot assume that p_{iA}/p_i is approximately constant, but will use one of the bounds given by (2.2.5). We shall calculate the approximate distribution of the D.S.N. under the null hypothesis, and will use the formula

$$P_m \leq \sum_{i=1}^m p_i \leq P_m + \sum_{i=1}^{m-1} p_i \eta_{im} + \alpha \max_{i, m} \delta_{im}.$$

We find that $\max_{i,m} \delta_{im}$ is very nearly equal to 1 for a number of combinations of i and m and therefore we can simplify our calculations by calculating the bounds from the formula

$$P_m \leq \sum_{i=1}^m p_i \leq P_m + \sum_{i=1}^{m-1} p_i \eta_{im} + \alpha,$$

where

$$\eta_{im} = \int_{2.475(m-i)}^{20.700+2.475(m-i)} p(\chi_{m-i}^2) d\chi_{m-i}^2$$

(since $\sum_{j=i+1}^m x_j^2$ is distributed as $\chi_{m-i}^2 \sigma^2$).

The following table gives the approximate D.S.N. distribution, together with the upper and the lower bounds for the cumulative D.S.N. distribution from which the approximate distribution was obtained by averaging.

Table 4. Estimation of p_i for Example 3

Sample no. i	Lower bound L_i	Upper bound U_i	Mean of the bounds $\frac{1}{2}(L_i + U_i)$	Approximate probabilities $\hat{p}_i$
5	0.13722	0.13722	0.13722	0.13722
6	.39140	.41729	.40434	.26712
7	.56689	.61939	.59314	.18880
8	.69340	.75623	.72481	.13167
9	.77997	.84338	.81167	.08686
10	0.84448	0.90972	0.87360	0.06193
11	.88878	.94038	.91458	.04098
12	.91951	.96396	.94178	.02715
13	.94342	.98146	.96242	.02069
14	.95838	.99105	.97472	.01230
≥ 15	0.96991	—	1.00000	0.02528

The region A_m does not exist for values of m less than 5 and therefore decision to accept H_0 is not possible for these values of m ; hence the above calculations are for values of $m \geq 5$. The probability p_5 contains, in addition to the probability of reaching a decision at $m = 5$, the probability of reaching a decision to reject H_0 at $m = 1, 2, 3$ or 4 which is almost negligible.

4. BOUNDS FOR THE AVERAGE SAMPLE NUMBER (A.S.N.) OF A SEQUENTIAL TEST

If we write, abbreviating (2.2.4),

$$P_m \leq \sum_{i=1}^m p_i \leq P'_m,$$

then it is easy to show that, for sufficiently large N ,

$$N + 1 - \sum_{m=1}^N P'_m \leq \sum_{j=1}^N j p_j \leq N + 1 - \sum_{m=1}^N P_m. \quad (4.1)$$

As N tends to infinity the central term in these inequalities tends to the A.S.N. of the test. In Table 5 the limits for the A.S.N. so obtained are compared with the limits given by Wald for the problems stated in Examples 1 and 2.

The experimental value 8.8 of the A.S.N. was taken from Baker (1950) and is based on the results of 2002 tests. The other experimental value 5.65 is based on the unpublished results of the 698 tests carried out by Baker.

From this table we can see that the lower bound, given by (4.1), is sharper than that given by Wald, but that Wald gives the sharper upper bound. Hence, the lower bound in (4.1), and Wald's upper bound, give narrower limits than either system alone.

Table 5. Comparison of upper and the lower bounds for the A.S.N. obtained by using Wald's formula with those obtained using equation (4.1.1). Case $H_0: \theta = 0; H_1: \theta = 1$

Risks ...	$\alpha = \beta = 0.025$			$\alpha = \beta = 0.0758$			
	Wald	Equation (4.1.1)	Expt.	Wald	Equation (4.1.1)	Expt.	Exact
Lower bound	6.9	7.61	—	4.20	4.94	—	—
Approx. A.S.N.	7.2	8.62	8.8	4.24	5.81	5.65	5.87
Upper bound	8.9	9.64	—	6.10	7.20	—	—

5. CONCLUSION

Many of the results in this paper are valid for tests of composite hypothesis also. It is hoped to publish the results of similar investigations for composite hypothesis in the near future.

The author wishes to thank Dr H. O. Hartley and Dr N. L. Johnson for their valuable suggestions and guidance, while the author was working in the Department of Statistics, University College London, and also the referee for suggestions and criticisms.

REFERENCES

- ALBERT, G. E. (1954). On the computation of sampling characteristics of a general class of sequential decision problems. *Ann. Math. Statist.* **25**, 340-56.
- BAKER, A. G. (1950). Properties of some tests in sequential analysis. *Biometrika*, **37**, 334-46.
- BHATE, D. H. (1958). A note on the distribution of the successive sums of samples from an exponential population. *Bull. Calcutta Statist. Ass.* **8**, 13-19.
- DARLING, D. A. & SIEGERT, A. T. (1953). First passage problem for a continuous Markoff process. *Ann. Math. Statist.* **24**, 624-39.
- KAC, M. (1945). Random walk in the process of absorbing barriers. *Ann. Math. Statist.* **14**, 62-7.
- PAGE, E. S. (1954a). Monte Carlo solutions of some integral equations. *Proc. Camb. Phil. Soc.* **50**, 414-26.
- PAGE, E. S. (1954b). An improvement of Wald's approximations for some properties of sequential tests. *J.R. Statist. Soc. Ser. B*, **16**, 136-9.
- RAMCHANDRAN, G. (1954). Some contributions to sequential analysis. Ph.D. thesis, University of Madras.
- WALD, A. (1943). *Sequential Analysis of Statistical Data: Theory*. Report submitted to Applied Mathematics Panel; National Defence Research Committee.
- WALD, A. (1947). *Sequential Analysis*. New York: John Wiley and Sons.

TOURNAMENTS AND PAIRED COMPARISONS*

By H. A. DAVID

Virginia Polytechnic Institute and University of Melbourne

1. INTRODUCTION AND SUMMARY

In an experiment involving paired comparisons t 'stimuli' such as flavours or colours are presented in pairs to a panel of n judges. The basic experimental unit is the comparison of two stimuli by a single judge who must state which one he prefers. Such an individual paired comparison may be regarded as a game between two contestants—the stimuli—while the experiment as a whole is analogous to a tournament of t players. If all the possible $\frac{1}{2}t(t-1)$ comparisons are made by each of the n judges, this corresponds to a Round Robin tournament repeated n times. It will be convenient in this paper to use the language of tournaments rather than that of paired comparisons.

The analogy between a paired comparison experiment and a Round Robin tournament, first pointed out by Kendall (1955), suggests the investigation of another well-known way of running competitions, the Knock-out tournament. It is perhaps surprising that such tournaments do not appear to have been studied at all in the statistical literature, although this may be due to their inherent lack of balance†. However, while this objection is valid if a ranking of all players is desired, it loses much of its force when the aim is to pick the best player. In §2 some properties of a simple (i.e. unrepeated) Knock-out tournament are investigated; in particular, an expression is found for the probability with which the best player wins such a tournament, various assumptions about the strengths of the players being made.

Round Robin tournaments (in effect) have received considerable attention both from statisticians and psychometricians. One problem is the enumeration of the possible outcomes, expressed as scores, of a tournament of a given size and the determination of the frequencies of these outcomes. The matter has been investigated by Bradley & Terry (1952) and Bradley (1954) and is further considered in §3 and Table 1. Bradley & Terry have also proposed a model to represent the strengths π_i ($i = 1, 2, \dots, t$; $\sum \pi_i = 1$) of the players which allows the π_i to be estimated from the players' scores. A different model had been suggested earlier by Thurstone (1927). Tests of the over-all equality of scores may be based on these methods or on combinatorial arguments leading to a test criterion involving the sum of squares of the scores (Durbin, 1951). For a simple Round Robin tournament the latter approach is equivalent to the 'circular triad test' of Kendall & Babington Smith (1940).

However, these approaches do not provide an answer to such questions as what constitutes a significant difference in the scores of two specified players. To deal with problems of this type, the joint probability distribution of the scores of any s ($s \leq t$) players in a simple Round Robin tournament is found in §3.1 on the hypothesis, H_0 , that the players are all of equal strength. This result is used to obtain tests of significance based on what are essentially upper percentage points of (i) the score of a specified player, (ii) the difference in the scores of two specified players, and (iii) the top score in the tournament.

Some aspects of repeated tournaments are discussed in §4.

* Research supported, in part, by the Office of Ordnance Research, U.S. Army.

† [See, however, a paper by Rita J. Maurice (*Biometrika* (1958), 45, 581-6) published since this paper was written. Ed.]

2. KNOCK-OUT TOURNAMENTS

We suppose that the number t of players has been reduced, by preliminary play-offs or otherwise, to a power of 2, say $t = 2^p$. The tournament will then be completed in p rounds. If all contestants are of equal strength, each has probability $t^{-1} = 2^{-p}$ of winning the tournament; or, more briefly, each has *winning probability* 2^{-p} . Likewise, if a particular player beats any one opponent with probability π , he will win the tournament with probability π^p .

In more complicated cases the winning probability of a specified player will evidently depend on the initial tournament draw, which we take to be strictly random. Let us call the t contestants $C_1, C_2, \dots, C_t$ and write

$$C_i \xrightarrow{\pi} C_j \quad (i, j = 1, 2, \dots, t; i \neq j),$$

to express that C_i defeats C_j with probability π . Let $P_i^{(r)}$ ($r = 1, 2, \dots, p$) denote the probability that C_i reaches the r th round of the tournament *and* wins his game in this round, so that $P_i^{(p)}$ is his probability of winning the tournament. Finally, let $M_{ij}^{(r)}$ be the probability that C_i meets C_j in the r th round and M_{ij} be the probability of their meeting in the course of the tournament. Clearly

$$M_{ij} = \sum_{r=1}^p M_{ij}^{(r)}. \quad (2.1)$$

Suppose now that

$$C_1 \xrightarrow{\pi_1} C_3, C_4, \dots, C_t,$$

$$C_2 \xrightarrow{\pi_2} C_3, C_4, \dots, C_t,$$

$$C_1 \xrightarrow{\pi_{12}} C_2.$$

Then C_1 's probability of winning the tournament is given by

$$P_1^{(p)} = M_{12} \pi_{12} \pi_1^{p-1} + (1 - M_{12}) \pi_1^p. \quad (2.2)$$

If the draw is known, M_{12} and hence $P_1^{(p)}$ can be evaluated. Instead of considering this, we shall obtain expressions for $\bar{M}_{12}$ and $\bar{P}_1^{(p)}$, the values of M_{12} and $P_1^{(p)}$ *before* the draw. These prior probabilities will still be linked by (2.2). We have

$$\bar{M}_{12}^{(1)} = 1/(t-1).$$

If C_1 and C_2 are to meet in the second round, in which there will be $\frac{1}{2}t$ contestants, they must both be successful in the first round, so that

$$\begin{aligned} \bar{M}_{12}^{(2)} &= (1 - \bar{M}_{12}^{(1)}) \pi_1 \pi_2 / (\tfrac{1}{2}t - 1) \\ &= 2\pi_1 \pi_2 / (t - 1). \end{aligned}$$

Continuing the argument, we find

$$\bar{M}_{12}^{(r)} = (2\pi_1 \pi_2)^{r-1} / (t - 1),$$

and hence from (2.1)

$$\bar{M}_{12} = \sum_{r=1}^p (2\pi_1 \pi_2)^{r-1} / (t - 1).$$

It is interesting to note that the $\bar{M}_{ij}^{(r)}$ depend on π_1 and π_2 only through the product $\pi_1 \pi_2$. We see also that the probability of a final between C_1 and C_2 is

$$\bar{M}_{12}^{(p)} = (2\pi_1 \pi_2)^{p-1} / (t - 1).$$

If for the moment we imagine C_1 and C_2 to be 'seeded' players placed in opposite halves of an otherwise random draw, this probability would become

$$(\pi_1 \pi_2)^{p-1} \div 2\bar{M}_{12}^{(p)}.$$

We turn now to the general case where C_i defeats C_j with probability π_{ij} , the only restriction on the π_{ij} being $\pi_{ji} = 1 - \pi_{ij}$ (due to the absence of ties). It should be noted that the possibility

$$\pi_{ij} > \tfrac{1}{2}, \quad \pi_{jk} > \tfrac{1}{2}, \quad \pi_{ki} > \tfrac{1}{2}$$

is not excluded (compare Kendall & Babington Smith, 1940). This situation can occur when the scale of playing strength is not linear, in which case we define the best player as the one with the largest expected score. The models of Thurstone and Bradley & Terry assume a linear scale corresponding to the choice of suitable expressions for the π_{ij} . As we are not concerned with estimation, they need no separate treatment here.

To illustrate the general case take $t = 8$ and suppose (without loss of generality) that $C_1, \dots, C_8$ have been drawn in that order. Then we have, for example,

$$P_1^{(1)} = \pi_{12}, P_2^{(1)} = \pi_{21}, P_3^{(1)} = \pi_{34}, \dots, \quad (2.3)$$

$$P_1^{(2)} = P_1^{(1)}(\pi_{13}P_3^{(1)} + \pi_{14}P_4^{(1)}) = \pi_{12}(\pi_{13}\pi_{34} + \pi_{14}\pi_{43}), \quad (2.4)$$

$$P_1^{(3)} = P_1^{(2)}(\pi_{15}P_5^{(2)} + \pi_{16}P_6^{(2)} + \pi_{17}P_7^{(2)} + \pi_{18}P_8^{(2)}). \quad (2.5)$$

The corresponding probabilities before the draw can be obtained by averaging over all possible draws. Let $P_1^{(2)}(i, j, k)$ denote the joint probability that C_1 beats C_i in the first round and the winner of $C_j v. C_k$ in the second round; then $P_1^{(2)}(2, 3, 4)$ is the expression of (2.4). Likewise, we may (for $t = 8$) write $P_1^{(3)}$ of (2.5) more fully as $P_1^{(3)}(2, 3, 4)$. It follows that

$$\bar{P}_1^{(1)} = \frac{1}{7} \sum_{i=2}^8 \pi_{1i}, \quad (2.6)$$

$$\bar{P}_1^{(2)} = \frac{1}{105} \sum_i \sum_j \sum_{\substack{k \\ 1 \neq i \neq j \neq k}} P_1^{(2)}(i, j, k), \quad (2.7)$$

$$\bar{P}_1^{(3)} = \frac{1}{105} \sum_i \sum_j \sum_{\substack{k \\ 1 \neq i \neq j \neq k}} P_1^{(3)}(i, j, k).$$

For $t > 8$ the number of terms on the right-hand side may become formidable. Thus $\bar{P}_1^{(r)}$ is the mean of

$$(t-1) \binom{t-2}{2} \binom{t-4}{4} \dots \binom{t-2r+1}{2r-1} \text{ terms.}$$

Example. In a tournament of 8 players, let C_i 's 'strength' in his encounter with C_j be represented by

$$x_{ij} = \xi_i + v_{ij},$$

where ξ_i is C_i 's expected strength and the error variate v_{ij} is $N(0, \frac{1}{2}\sigma^2)$, the 56 v_{ij} 's being mutually independent (compare Mosteller, 1951).

Then

$$\pi_{ij} = \Pr(\xi_j + v_{ji} < \xi_i + v_{ij}) = \int_{-\infty}^{(\xi_i - \xi_j)/\sigma} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}u^2} du.$$

We now take ξ_i to be the expected value of the i th largest unit normal deviate in samples of 8, so that the players are arranged in decreasing order of strength with C_1 the strongest. The ξ_i have been tabulated by Teichroew (1956). For $\sigma = 1$ the following array of π_{ij} 's results:

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
C_1	—	0.7161	0.8291	0.8982	0.9425	0.9710	0.9886	0.9978
C_2	0.2839	—	0.6478	0.7579	0.8425	0.9074	0.9558	0.9886
C_3	0.1709	0.3522	—	0.6256	0.7341	0.8278	0.9074	0.9710
C_4	0.1018	0.2421	0.3744	—	0.6198	0.7341	0.8425	0.9425
C_5	0.0575	0.1575	0.2659	0.3802	—	0.6256	0.7579	0.8982
C_6	0.0290	0.0926	0.1722	0.2659	0.3744	—	0.6478	0.8291
C_7	0.0114	0.0442	0.0926	0.1575	0.2421	0.3522	—	0.7161
C_8	0.0022	0.0114	0.0290	0.0575	0.1018	0.1709	0.2839	—

It will be seen that due to the symmetry of the normal distribution

Applying (2.6) and (2.7) we obtain $\pi_{ij} = \pi_{9-j, 9-i}$.

$$\begin{aligned}\bar{P}_i^{(1)} &= 0.906, & 0.769, & 0.656, & 0.551, & 0.449, & 0.344, & 0.231, & 0.094; \\ \bar{P}_i^{(2)} &= 0.781, & 0.494, & 0.319, & 0.200, & 0.117, & 0.060, & 0.024, & 0.004.\end{aligned}$$

It is interesting to compare these probabilities before the draw with those in a draw in which considerable seeding has been used, the first round games being $C_1 v. C_8, C_5 v. C_4, C_3 v. C_6, C_7 v. C_2$. From (2.3) to (2.5) we find

$$\begin{aligned}P_i^{(1)} &= 0.998, & 0.956, & 0.828, & 0.620, & 0.380, & 0.172, & 0.044, & 0.002; \\ P_i^{(2)} &= 0.913, & 0.662, & 0.312, & 0.064, & 0.023, & 0.020, & 0.006, & 0.002; \\ P_i^{(3)} &= 0.692, & 0.216, & 0.066, & 0.019, & 0.005, & 0.001, & 0.000, & 0.000.\end{aligned}$$

3. ROUND ROBIN TOURNAMENTS

The results of a Round Robin tournament can be presented fully in the familiar two-way table of 1's and 0's, with the scores (i.e. number of wins) of each player listed in a column of totals. As a summary of the scores, we may write $(a_1 a_2 \dots a_t)$, meaning thereby that C_1 has won a_1 games, C_2 has won a_2 games, and so on; clearly $\sum a_i = \frac{1}{2}t(t-1)$. If the a 's differ greatly, a real difference between at least some of the players is indicated. Thus a Round Robin tournament, unlike a Knock-out tournament—or a ranking—provides the possibility of a test of the null hypothesis, H_0 , that all players are of equal strength even when there has been no replication. From this point of view the identity of the players is of no consequence, and we may replace $(a_1 a_2 \dots a_t)$ by the partition $[x_1^r x_2^r \dots x_m^r]$, where the x 's are simply a re-arrangement of the a 's in decreasing order of magnitude, and r_1 is the number of occurrences of the largest score x_1 , etc.* It follows that

$$\sum_{u=1}^m r_u = t, \quad \sum_{u=1}^m r_u x_u = \frac{1}{2}t(t-1).$$

One test of H_0 is the 'circular triad test' due to Kendall & Babington Smith (1940); another has been given by Bradley & Terry (1952) and is based on their model referred to earlier. In both cases the test-criterion is a symmetric function of the scores, and its null distribution is easily deduced from a knowledge of the different permissible partitions of $\frac{1}{2}t(t-1)$ together with their frequencies of occurrence. Such basic tables have been constructed by Bradley & Terry (1952), and Bradley (1954) for $t \leq 5$. These authors also show how results for several complete repetitions may be built up from those for $n = 1$ and their tables cover many such cases. In Table 1 we have extended their results for $n = 1$ to $t \leq 8$, using a score of 1 for a win and 0 for a loss, rather than 1 for a win and 2 for a loss.

A possible approach to the enumeration problem involved is to consider the generating function

$$G(t) = \prod_{i < j} (b_i + b_j), \quad (3.1)$$

a product of $\frac{1}{2}t(t-1)$ terms corresponding to the games played. The expansion of $G(t)$ contains $2^{\frac{1}{2}t(t-1)}$ terms, the exponents in each of which correspond to a possible set of scores. For example,

$$G(3) = (b_1 + b_2)(b_1 + b_3)(b_2 + b_3) = \sum b_1^2 b_2 + 2b_1 b_2 b_3, \quad (3.2)$$

* We depart slightly from standard notation in showing zeros explicitly.

Table 1. *Partitions of scores and their frequencies in a Round Robin tournament of t players*

[illegible]

Table 1 (cont.)

$t = 8$ (cont.)		$t = 8$ (cont.)		$t = 8$ (cont.)		$t = 8$ (cont.)	
$[6^3 3^2 21^2]$	26,880	$[6^2 43^2 2^3]$	1,464,960	$[6543^4 1]$	3,790,080	$[5^3 43^2 21]$	7,889,280
$6^3 3^2 3^1$	31,360	$6^2 3^5 1$	142,464	$6543^3 2^2$	12,230,400	$5^3 43^2 3$	4,282,880
$6^3 2^5$	2,688	$6^2 3^4 2^2$	577,920	$653^5 2$	1,876,224	$5^3 4^1 1$	1,025,920
$6^2 5^2 3210$	40,320	$65^3 4210$	94,080			$5^3 3^3 2^2$	3,306,240
$6^2 5^2 31^3$	13,440			$64^5 20$	142,464	$5^2 4^4 20$	577,920
		$65^3 41^3$	31,360	$64^5 1^2$	142,464	$5^2 4^4 1^2$	577,920
$6^2 5^2 2^3 0$	13,440	$65^3 3^2 10$	94,080	$64^4 3^2 0$	577,920	$5^2 4^3 3^2 0$	1,868,160
$6^2 5^2 2^2 1^2$	40,320	$65^3 3^2 2^0$	188,160	$64^4 321$	3,790,080		
$6^2 5^4 2^2 10$	80,640	$65^3 321^2$	376,320	$64^4 2^3$	1,025,920	$5^2 4^3 321$	12,230,400
$6^2 5^4 2^1 3$	26,880	$65^3 2^3 1$	219,520			$5^2 4^3 2^3$	3,306,240
$6^2 5^4 3^2 10$	161,280			$64^3 3^3 1$	3,924,480	$5^2 4^3 2^3 1$	9,461,760
		$65^2 4^2 310$	483,840	$64^3 3^2 2^2$	9,461,760	$5^2 4^3 2^3 2^2$	22,780,800
$6^2 5^4 32^2 0$	322,560	$65^2 4^2 2^2 0$	483,840	$64^2 3^4 2$	7,203,840	$5^2 4^3 4^2$	11,518,080
$6^2 5^4 321^2$	645,120	$65^2 4^2 21^2$	967,680	643^6	725,760		
$6^2 5^4 2^3 1$	376,320	$65^2 4^3 2^2 0$	1,653,120	$5^5 210$	8,064	$5^2 3^6$	577,920
$6^2 5^3 3^2 0$	188,160	$65^2 4^3 2^1 2$	1,653,120			$54^5 30$	577,920
$6^2 5^3 31^2$	188,160			$5^5 1^3$	2,688	$54^5 21$	1,876,224
		$65^2 4^3 2^2 1$	5,644,800	$5^4 4310$	134,400	$54^4 3^2 1$	7,203,840
$6^2 5^3 2^2 2^1$	967,680	$65^2 4^2 4^1$	779,520	$5^4 42^2 0$	134,400	$54^4 32^2$	11,518,080
$6^2 5^3 2^4$	268,800	$65^2 3^4 0$	228,480	$5^4 421^2$	268,800		
$6^2 4^3 310$	94,080	$65^2 3^3 21$	3,037,440	$5^4 3^2 20$	228,480	$54^3 3^3 2$	23,157,120
$6^2 4^3 2^2 0$	94,080	$65^2 3^2 2^3$	2,486,400			$54^3 3^5$	5,201,280
$6^2 4^3 21^2$	188,160			$5^4 3^2 1^2$	228,480	47^0	21,120
		$65^4 410$	134,400	$5^4 32^2 1$	779,520	$4^6 31$	725,760
$6^2 4^2 3^2 20$	483,840	$65^4 3^3 20$	1,787,520	$5^4 2^4$	107,520	$4^6 2^2$	577,920
$6^2 4^2 3^2 1^2$	483,840	$65^4 3^3 1^2$	1,787,520	$5^3 4^3 10$	147,840		
$6^2 4^2 3^2 2^1$	1,653,120	$65^4 3^2 2^1$	3,037,440	$5^3 4^2 320$	1,464,960	$4^5 3^2 2$	5,201,280
$6^2 4^2 2^4$	228,480	$65^4 2^3 30$	1,464,960			$4^4 3^4$	3,230,080
$6^2 4^3 40$	134,400			$5^3 4^2 31^2$	1,464,960		
		$65^4 2^3 2^2 1$	14,515,200	$5^3 4^2 2^2 1$	2,486,400	Total	268,435,456
$6^2 4^3 3^2 21$	1,787,520	$65^4 2^3 2^3$	7,889,280	$5^3 4^3 30$	797,440		

the sum being over 6 terms. This identity may be interpreted as follows. In a tournament of 3 players there are 6 outcomes of type $[210]$ where one player wins both his games, a second wins one game and the third wins none; and 2 outcomes of type $[1^3]$ where each player wins one game. $G(t)$ is symmetric in the a 's and may be expanded as in (3.2) as a sum of monomial symmetric functions. Table 1 has been constructed essentially by noting that

$$G(t+1) = G(t) \prod_{i=1}^t (b_{t+1} + b_i) = G(t) [b_{t+1}^t + (1^2) b_{t+1}^{t-1} + (1^2) b_{t+1}^{t-2} + \dots + (1^t)],$$

where

$$(1^s) = \sum b_1 b_2 \dots b_s, \quad (s = 1, 2, \dots, t).$$

Three generalizations of (3.1) are perhaps worth pointing out, although they do not necessarily provide an easy solution to the problems involved.

(a) If C_i and C_j play n_{ij} games instead of merely one, the appropriate generating function is

$$\prod_{i < j} (b_i + b_j)^{n_{ij}}.$$

(b) If $C_i \xrightarrow{\pi_{ij}} C_j$ for all $i \neq j$ then $G(t)$ can be replaced by

$$\prod_{i < j} (b_i \pi_{ij} + b_j \pi_{ji})$$

to generate the probabilities of the various partitions.

(c) Suppose that instead of being two-point (1 and 0) the scoring scale is $(r+1)$ -point $\{1, (r-1)/r, \dots, (1/r), 0\}$, with the total score in any one game still unity. Then

$$\prod_{i < j} (b_i + b_i^{(r-1)/r} b_j^{1/r} + b_i^{(r-2)/r} b_j^{2/r} + \dots + b_j)$$

generates all possible outcomes. This includes the case where games may end in either a win, a loss or a tie ($r = 2$).

3.1. Significance tests

As already mentioned, several methods are available for testing the null hypothesis H_0 that all t players in a Round Robin tournament are of equal strength. These tests are exact only for small values of t , but approximations for large t have also been given. However, they are overall tests and do not provide an answer to the more detailed questions we may wish to ask. We consider the following questions:

(i) What score must a *specified* player reach to be regarded as significantly better than average?

(ii) What constitutes a significant difference in scores between two specified players?

(iii) On the basis only of the scores obtained, what constitutes a significantly high score?

In order to deal with these problems the joint distribution of scores, on H_0 , of s ($\leq t$) players is required. C_i 's score clearly has the binomial distribution

$$f(a_i) = \binom{t-1}{a_i} \frac{1}{2^{t-1}}.$$

To find the joint distribution of a_i and a_j , we note that these scores arise if C_i defeats C_j and wins $a_i - 1$ of his other $t - 2$ games, while C_j wins a_j ; and similarly if C_j defeats C_i . Thus

$$f(a_i, a_j) = \frac{1}{2^{2t-3}} \left[\binom{t-2}{a_i-1} \binom{t-2}{a_j} + \binom{t-2}{a_i} \binom{t-2}{a_j-1} \right]. \quad (3.3)$$

The argument may be extended to give the joint distribution of s players $C_i, C_j, \dots, C_k$ by a consideration of the $2^{\frac{1}{2}s(s-1)}$ outcomes of the $\frac{1}{2}s(s-1)$ games between these s players. If their scores in this sub-tournament are $a'_i, a'_j, \dots, a'_k$, their final scores will be $a_i, a_j, \dots, a_k$ with probability

$$f(a_i, a_j, \dots, a_k | a'_i, a'_j, \dots, a'_k) = \frac{1}{2^{s(t-s)}} \binom{t-s}{a_i-a'_i} \binom{t-s}{a_j-a'_j} \dots \binom{t-s}{a_k-a'_k}, \quad (3.4)$$

provided we use the convention

$$\binom{n}{r} = 0 \quad \text{for } r < 0 \quad \text{and} \quad r > n. \quad (3.5)$$

It follows that

$$f(a_i, a_j, \dots, a_k) = \frac{1}{2^{\frac{1}{2}s(2t-s-1)}} \sum \prod_{l=i,j,\dots,k} \binom{t-s}{a_l-a'_l},$$

where the sum is, in effect, over all outcomes of the sub-tournament compatible with the final scores.

We turn now to questions (i)–(iii) above.

(i) *Test of the score of a specified player*

Let C_1 be the player in whose performance we are interested. Since, on H_0 , a_1 is a binomial variate with parameters $t-1, \frac{1}{2}$, it is easy to test H_0 against the alternative that C_1 is above average strength (one-sided binomial test) or different from the average strength of the other players (two-sided binomial test). For example, at the 5% level of significance the

following are the smallest values of a_1 which are significantly higher than the expected value $\frac{1}{2}(t-1)$:

t	a_1
≤ 5	no significantly high score possible
6-8	$t-1$
9-11	$t-2$
> 11	smallest integer $> \frac{1}{2}[t + 1.645\sqrt{(t-1)}]$.

The large-sample result follows from the normal approximation to the binomial, with continuity correction.

Next, suppose that the players are not of equal strength, but that it is desired to test whether C_1 's average probability of success π_1 equals $\frac{1}{2}$. We have

$$\mathcal{E}(a_1) = (t-1)\pi_1,$$

$$\text{var } a_1 = \sum_{j=2}^t \pi_{1j}(1-\pi_{1j}) \leq \frac{1}{4}(t-1).$$

Since the variance of a_1 is a maximum under H_0 , we may conclude that the above test will be conservative if H_0 is not true. In fact, as

$$\text{var } a_1 \leq (t-1)\pi_1(1-\pi_1)$$

we can also test the hypothesis that π_1 has a specified value π_0 by referring a_1 to the binomial distribution with parameters $t-1$, π_0 .

(ii) *Test of the equality of scores of two specified players*

We shall find the probability distribution $p(d)$ of $d = a_1 - a_2$, assuming H_0 to hold. From (3.3)

$$p(d) = \sum_{a_1} f(a_2 + d, a_2) = \frac{1}{2^{2t-3}} \left[\sum \binom{t-2}{a_2+d-1} \binom{t-2}{a_2} + \sum \binom{t-2}{a_2+d} \binom{t-2}{a_2-1} \right],$$

where in view of the convention (3.4), the summations may be taken over all non-negative integers a_2 . Thus

$$p(d) = \frac{1}{2^{2t-3}} \left[\binom{2t-4}{t-d-1} + \binom{2t-4}{t-d-3} \right], \quad d = -(t-1), -(t-2), \dots, (t-1),$$

is a symmetrical distribution whose variance is $\frac{1}{2}t$. It can be shown that the characteristic function of d is

$$\phi(u) = (\cos \frac{1}{2}u)^{2t-4} \cos u,$$

and hence that $d/\sqrt{(\frac{1}{2}t)}$ tends to a unit normal distribution. Numerical work indicates this tendency to be rapid.

At the 5% level of significance, the smallest significant values of $|d|$ are as follows:

t	$ d $
≤ 4	no significant values of $ d $ possible
5-6	4
7-10	5
> 10	smallest integer $> 1.96\sqrt{(\frac{1}{2}t)} + 0.5$.

While this approach tests equality of the two scores under H_0 it will also provide a conservative test of the null hypothesis

$$H'_0: \pi_1 = \pi_2 \quad \text{against the alternative } \pi_1 \neq \pi_2;$$

for, in the general case, we find

$$\text{var } d = 4\pi_{12}(1-\pi_{12}) + \sum_{j=3}^t [\pi_{1j}(1-\pi_{1j}) + \pi_{2j}(1-\pi_{2j})],$$

which is a maximum under H_0 .

(iii) *Test of the top score*

The test developed in this section is essentially different from those in (i) and (ii) in that it is not based on the scores of *specified* players, but on the score x_1 of the winning player (or players). We obtain, on H_0 , the probability of large values of x_1 from which upper percentage points may be constructed. If in a particular tournament the appropriate upper percentage point is attained or exceeded by one or more players, H_0 is rejected and, as in the analogous problem of testing for 'outliers', these players may be regarded as 'superior'.

It is easy to see that for $t \leq 8$ the probability distribution of x_1 could be built up from Table 1. We shall employ a different approach which is useful also for larger values of t . Since x_1 is the largest of the scores $a_1, a_2, \dots, a_t$, the method of inclusion and exclusion gives (see, for example, David, 1956)

$$\Pr(x_1 \geq t-r) = \sum_{s=1}^t (-1)^{s-1} \binom{t}{s} \Pr(a_1 \geq t-r, a_2 \geq t-r, \dots, a_s \geq t-r), \quad (3.6)$$

where r is any integer from 1 to t . For small values of r , the later terms in the sum will be zero, since only a few players can have a large score. We illustrate the method on the evaluation of the typical term

$$\Pr(a_1 \geq t-3, a_2 \geq t-3, a_3 \geq t-3, a_4 \geq t-3).$$

Consider the sub-tournament played by C_1, C_2, C_3 and C_4 . From Table 1 we see that the possible outcomes are

$$[3210], [31^3], [2^30], [2^21^2]$$

with respective frequencies

$$24, 8, 8, 24.$$

It is, however, clear that only $[31^3]$ and $[2^21^2]$ can result in final scores all greater than $t-4$. The possible outcomes (ignoring the order of C_1 to C_4), and the frequencies with which they can arise from $[31^3]$ and $[2^21^2]$, are as follows:

				$[31^3]$	$[2^21^2]$
				1	0
$t-1,$	$t-3,$	$t-3,$	$t-3$	0	1
$t-2,$	$t-2,$	$t-3,$	$t-3$	$t-4$	$2(t-4)$
$t-2,$	$t-3,$	$t-3,$	$t-3$	$\frac{1}{2}(t-4)(t-5)$	$(t-4)^2$
$t-3,$	$t-3,$	$t-3,$	$t-3$		

This gives a total frequency of $28t^2 - 172t + 272$, which can be converted into the required probability on division by $2^{2(t-5)}$, the factor of equation (3.5). In this way we obtain from (3.6)

$$\Pr(x_1 = t-1) = \frac{t}{2^{t-1}},$$

$$\Pr(x_1 \geq t-2) = \frac{t^2}{2^{t-1}} - \frac{\binom{t}{2}(t-1)}{2^{2t-4}} + \frac{\binom{t}{3}}{2^{3t-7}},$$

$$\begin{aligned} \Pr(x_1 \geq t-3) = & \frac{t}{2^t}(t^2 - t + 2) - \frac{\binom{t}{2}}{2^{2t-3}}(t^3 - 4t^2 + 7t - 4) \\ & + \frac{\binom{t}{3}}{2^{3t-6}}(5t^3 - 33t^2 + 78t - 64) - \frac{\binom{t}{4}}{2^{4t-12}}(7t^2 - 43t + 68) + \frac{3\binom{t}{5}}{2^{5t-18}}. \end{aligned}$$

These three probabilities are as follows for $t \leq 20$:

t	$\Pr(x_1 \geq t-1)$	$\Pr(x_1 \geq t-2)$	$\Pr(x_1 \geq t-3)$
3	0.7500	1.0000	1.0000
4	0.5000	1.0000	1.0000
5	0.3125	0.9766	1.0000
6	0.1875	0.8418	1.0000
7	0.1094	0.6447	0.9987
8	0.0625	0.4526	0.9637
9	0.0352	0.2989	0.8536
10	0.0195	0.1891	0.6868
11	0.0107	0.1161	0.5093
12	0.0059	0.0696	0.3548
13	0.0032	0.0410	0.2360
14	0.0017	0.0238	0.1517
15	0.0009	0.0137	0.0950
16	0.0005	0.0078	0.0584
17	0.0003	0.0044	0.0353
18	0.0001	0.0025	0.0211
19	0.0001	0.0014	0.0125
20	0.0000	0.0008	0.0073

We see from this table that an unbroken sequence of wins is not significant, at the 5 % level, until $t = 9$ in the case of a one-sided test and $t = 10$ for a two-sided test. For a one-sided test a score of $t - 2$ is significantly high for $t \geq 13$, and a score of $t - 3$ for $t \geq 17$.

4. DISCUSSION

Some properties of Knock-out and Round Robin tournaments have been presented. We now make a few remarks about their respective merits and drawbacks.

The attraction of Knock-out tournaments is, at least in part, that a large number of players can be handled quickly. With $t = 2^p$ competitors a decision is reached after p rounds and $t - 1$ games as against $t - 1$ rounds and $\frac{1}{2}t(t - 1)$ games required for a Round Robin tournament if no play-offs are necessary. A natural objection to Knock-out tournaments is that they are too unreliable. This is often true in the simple cases discussed in § 2, but in view of the relative brevity of such tournaments it may well be practicable to repeat them. The simplest way of doing this is to start with a fresh random draw for each repetition and to declare as winner the player with the largest total number of wins, possibly after a play-off. Alternatively, the initial draws of each repetition could be arranged to satisfy certain symmetry requirements. We do not recommend either of these approaches if the aim of the tournament is to rank all players in order of merit. For this purpose a simple Round Robin tournament will be preferable to $\frac{1}{2}t$ repetitions of a Knock-out tournament, although both require the same (minimum) number of games. However, the situation will be very different when we are primarily concerned with picking the best player, since the repeated Knock-out tournament will tend to include a larger number of encounters between the stronger players. Some work along these lines is in progress. There are also possibilities, for both types of

tournaments, of some sequential procedure whereby the number of contestants is progressively reduced.

In this paper attention has been concentrated on tournaments without repetition of any kind. When comparing a number of treatments in a paired comparison experiment, we will often require some repetition, possibly by different judges. If repeated Round Robin tournaments are impracticably large, the incomplete block type designs put forward by Kendall (1955) and Bose (1956) are helpful. However, to answer the type of question considered in § 3, generalizations of tests there given as well as new tests are required. It is hoped to deal with some of these issues in a future publication.

I should like to thank Dr R. A. Bradley for some helpful discussions and Dr I. R. Savage for a valuable comment. My thanks are also due to Miss Betty Laby and Miss G. E. Jacobs of the University of Melbourne and to Mr A. Romano of the Virginia Polytechnic Institute for assistance with the computations.

REFERENCES

- BOSE, R. C. (1956). Paired comparison designs for testing concordance between judges. *Biometrika*, **43**, 113-21.
- BRADLEY, R. A. (1954). Rank analysis of incomplete block designs. II. Additional tables for the method of paired comparisons. *Biometrika*, **41**, 502-37.
- BRADLEY, R. A. & TERRY, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, **39**, 324-45.
- DAVID, H. A. (1956). On the application to statistics of an elementary theorem in probability. *Biometrika*, **43**, 85-91.
- DURBIN, J. (1951). Incomplete blocks in ranking experiments. *Brit. J. Psych. (Statistics Section)*, **4**, 85-90.
- KENDALL, M. G. (1955). Further contributions to the theory of paired comparisons. *Biometrika*, **11**, 43-62.
- KENDALL, M. G. & BABINGTON SMITH, B. (1940). On the method of paired comparisons. I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, **16**, 3-9.
- MOSTELLER, F. (1951). Remarks on the method of paired comparisons. *Psychometrika*, **16**, 3-9.
- TEICHROEW, D. (1956). Tables of expected values of order statistics for samples of size twenty and less from the normal distribution. *Ann. Math. Statist.* **27**, 410-26.
- THURSTONE, L. L. (1927). Psychophysical analysis. *Amer. J. Psychol.* **38**, 368-89.

ESTIMATION OF THE NORMAL POPULATION PARAMETERS GIVEN A SINGLY CENSORED SAMPLE

By J. G. SAW

University College London

1. INTRODUCTION

The problem of estimating the parameters of a normal population given an ordered sample $x_1 < x_2 \dots < x_n$ censored above x_r has been treated by several authors.

Gupta (1952) has investigated the maximum likelihood solutions and the most efficient linear unbiased estimators for μ and σ , the location and scale parameters. The simultaneous equations of the maximum likelihood solution are not immediately soluble. In his paper Gupta provides a table which, if enlarged to give finer argument intervals, would make the solution comparatively rapid. No information, however, seems to be available as to the extent to which the resulting estimates of μ and σ are biased; this is likely to be of importance in small samples. The estimators proposed below, while based on the same statistics as those involved in the maximum likelihood equations, namely x_r , and the first and second moments of the observations about x_r , combine them in such a way that the estimates of μ and σ^2 are unbiased, however small the sample.

The method of most efficient linear unbiased estimators, whilst more easily handled computationally, suffers from two defects. In order to obtain the system of weights, it is necessary to invert a matrix of order equal to the number (r) of observations available. These weights differ for every set of values taken by r and n , thus requiring extensive tabling and arduous preliminary computations. Further, and this is a more serious difficulty, the weights are functions of $\mathcal{E}(y_i)$, $\mathcal{E}(y_i^2)$ and $\mathcal{E}(y_i y_j)$, where $y_1 < y_2 \dots < y_n$ is an ordered sample from the standard normal population. At present these are only tabled for sample sizes $n \leq 20$, (Sarhan & Greenberg, 1956).

Gupta has overcome the labour of the matrix inversion by proposing a modified linear estimator, the weights being simple functions of $\mathcal{E}(y_i)$ only. These modified estimators are reasonably efficient but we are again limited in the use of them by the fact that, as yet, the values of $\mathcal{E}(y_i)$ are not tabulated with sufficient accuracy for $n > 20$.

2. ESTIMATORS SYMMETRICAL IN $x_1, x_2, \dots, x_{r-1}$

If we take

$$\mu^* = \epsilon \bar{x}_{r-1} + (1 - \epsilon) x_r, \quad \text{with} \quad \bar{x}_{r-1} = \frac{1}{r-1} \sum_{i=1}^{r-1} x_i,$$

$$\eta^* = \alpha \sum_{i=1}^{r-1} (x_i - x_r)^2 + \beta \left[\sum_{i=1}^{r-1} (x_i - x_r) \right]^2$$

as estimators of μ and σ^2 , respectively, the ϵ being chosen to make μ^* unbiased and α and β being chosen to make η^* unbiased and of minimum variance, then providing these estimators are reasonably efficient, they will have immediate advantage owing to the small number of weights which have to be established. In fact it is found that the large sample efficiency of μ^* varies between 0.948 when $r/(n+1) = 0.20$ and 0.993 when $r/(n+1) = 0.80$.

Further, μ^* has a smaller variance than the corresponding modified estimator due to Gupta. The asymptotic efficiency of η^* is unity for all values of $r/(n+1)$.

Apart from the fact that they occur in the maximum likelihood equations, the reason for considering such statistics as $\bar{x}_{r-1}$, $\sum_{i=1}^{r-1} (x_i - x_r)^2$ and $\left[\sum_{i=1}^{r-1} (x_i - x_r) \right]^2$ is that the values $x_1, \dots, x_{r-1}$ enter symmetrically. Thus we may regard $x_1, \dots, x_{r-1}$ as an ordered sample from the population conditional on x_r , with density function

$$P(x | x_r) = \frac{1}{\sigma \sqrt{(2\pi)}} \exp \left\{ -\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right\} \bigg/ \int_{-\infty}^{(x_r - \mu)/\sigma} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}u^2} du \quad (-\infty < x < x_r).$$

However, for the purposes of this work, since μ^* and η^* are symmetrical in $x_1, \dots, x_{r-1}$, we need no longer regard them as ordered. In effect therefore we no longer need the individual values of $\mathcal{E}(y_i)$, $\mathcal{E}(y_i^2)$ and $\mathcal{E}(y_i y_j)$ mentioned in the introduction.

In a previous paper by the author (1958) it was shown how it was possible to obtain moments and product moments of $\bar{x}_{r-1}$ and x_r in terms of a linear sum of integrals of the type

$$\psi(p_r, n; a, b) = \frac{1}{B(r, n-r+1)} \int_{-\infty}^{\infty} Z(u) [F(u)]^{r-1} [1-F(u)]^{n-r} \left[\frac{Z(u)}{F(u)} \right]^a u^b du,$$

where $Z(u) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}u^2}$, $F(u) = \int_{-\infty}^u Z(t) dt$ and $p_r = r/(n+1)$,

and also how it was possible to expand these integrals as a power series in $1/(n+2)$; tables were given for $a+b \leq 4$, $p_r = 0.50(0.05)0.80$. We may treat moments and product moments of $\sum (x_i - x_r)^2$ and $[\sum (x_i - x_r)]^2$ in a similar manner. Ultimately, therefore, ϵ , α and β may be expressed as power series in $1/(n+2)$ though they are given at the end of the paper as power series in $1/(n+1)$, since this form is thought to converge more rapidly. In addition, it is possible using Sarhan & Greenberg's tables to give exact values for ϵ and $\text{var}(\mu^*)$ for $1 < r < n \leq 20$.

3. ESTIMATION OF μ

Writing

$$\mu'_{10} = \mathcal{E}[(\bar{x}_{r-1} - \mu) | \sigma], \quad \mu'_{01} = \mathcal{E}[(x_r - \mu) | \sigma], \quad \mu_{ij} = \sigma^{-(i+j)} \mathcal{E}[(\bar{x}_{r-1} - \mathcal{E}(\bar{x}_{r-1}))^i (x_r - \mathcal{E}(x_r))^j]$$

we see that if $\mu^* = \epsilon \bar{x}_{r-1} + (1-\epsilon)x_r$ is to be an unbiased estimator for μ , then we must have

$$\epsilon = \mu'_{01} / (\mu'_{01} - \mu'_{10}).$$

With this value of ϵ

$$\text{var}(\mu^*) = \sigma^2 (\mu_{01}^{\prime 2} \mu_{20} - 2\mu'_{01} \mu'_{10} \mu_{11} + \mu_{10}^{\prime 2} \mu_{02}) / (\mu'_{01} - \mu'_{10})^2.$$

Exact values of $\epsilon(r, n)$ and $\text{var}(\mu^*)$ are given in Table 1 for sample sizes $n \leq 20$.

For use with sample sizes greater than twenty, the expansion

$$\epsilon(r, n) = \sum_{i=0}^5 E_i(p_r) (n+1)^{-i} + O(n+1)^{-6} \quad \text{with} \quad p_r = r/(n+1)$$

was obtained and will be found to give an excellent approximation to the true values at $n = 19$ (correct to the eighth place of decimals in most cases). The weight ϵ is a function of both r and n ; the complete form $\epsilon(r, n)$ has been used above in view of a relationship between $\epsilon(r, n)$ and $\epsilon(n-r+1, n)$ which is given below.

Similarly, we obtain an expansion for the variance of μ^* of the form

$$\text{var} \left(\frac{\mu^*}{\sigma} \right) \Big|_{(r,n)} = \sum_{i=1}^5 V_i(p_r) (n+1)^{-i} + O(n+1)^{-6}.$$

The quantities $E_i(p_r)$ and $V_i(p_r)$, which are functions of the ratio $p_r = r/(n+1)$ and i only, are given in Table 2 for $p_r = 0.50$ (0.05) 0.80. Values to cover the interval $0.20 \leq p_r < 0.50$ may be obtained using the relations

$$\frac{r-1}{\epsilon(r,n)} + \frac{n-r}{\epsilon(n-r+1,n)} = n,$$

$$\text{var} \left(\frac{\mu^*}{\sigma} \right) \Big|_{(r,n)} = \frac{1}{n} + \frac{\text{var}(\mu^*/\sigma) \Big|_{(n-r+1,n)} - (1/n)}{\{1 - n\epsilon(n+1-r,n)/(n-r)\}}.$$

4. COMPARISON OF μ^* WITH OTHER ESTIMATORS

Gupta has tabled the large sample variance of $\hat{\mu}$, the maximum likelihood estimator of μ . Comparing these tables with the leading term $V_1(p_r)(n+1)^{-1}$, we obtain the asymptotic efficiency of μ^* for various values of p_r :

p_r	Asymptotic efficiency	p_r	Asymptotic efficiency
0.20	0.9485	0.55	0.9711
0.25	0.9479	0.60	0.9764
0.30	0.9491	0.65	0.9813
0.35	0.9519	0.70	0.9857
0.40	0.9559	0.75	0.9896
0.45	0.9606	0.80	0.9929
0.50	0.9658		

We may also compare the efficiency of μ^* and μ_G (Gupta's modified estimator), relative to μ_L , the most efficient linear unbiased estimator. Thus for a sample of size ten we have

r	$\frac{\text{var } \mu_L}{\text{var } \mu^*}$	$\frac{\text{var } \mu_L}{\text{var } \mu_G}$
2	1.0000	1.0000
3	0.9819	0.9150
4	0.9756	0.8983
5	0.9771	0.9113
6	0.9825	0.9354
7	0.9890	0.9663
8	0.9946	0.9806
9	0.9985	0.9943

It should also be noticed that $\epsilon = 0$ whenever $p_r = r/(n+1) = \frac{1}{2}$. Thus if we are given the observations up to and including the median, we disregard all but the median value in forming an estimate of μ . Although this will at first sight seem a great waste of data, it will be observed from the first table in this section that the asymptotic efficiency of the median in this context is 96.58 %, whilst the ratio (variance μ_L)/(variance μ^*) takes the values:

$$0.9899 \quad \text{when } n = 5, r = 3; \quad 0.9843 \quad \text{when } n = 7, r = 4;$$

$$\text{and} \quad 0.9806 \quad \text{when } n = 9, r = 5.$$

Thus when only $\frac{1}{2}(n+1)$ of the observations are available, very little is lost by neglecting all but the median observation as far as the estimate of μ is concerned. In contrast, the asymptotic efficiency of the median relative to the complete sample of n is 63.66 %.

5. ESTIMATION OF σ^2

Defining

$$H'_{ij} = \mathcal{E} \left\{ \left[\sum_{l=1}^{r-1} (x_l - x_r)^2 \right]^i \left[\sum_{l=1}^{r-1} (x_l - x_r) \right]^{2j} \right\} \sigma^{-2(i+j)},$$

we see that

$$\eta^* = \alpha \sum_{i=1}^{r-1} (x_i - x_r)^2 + \beta \left[\sum_{i=1}^{r-1} (x_i - x_r) \right]^2$$

is unbiased for σ^2 and of minimum variance over all possible (α, β) if

$$\alpha = (H'_{10}H'_{02} - H'_{01}H'_{11}) / (H'_{10}{}^2H'_{02} + H'_{01}{}^2H'_{20} - 2H'_{10}H'_{01}H'_{11}),$$

$$\beta = (H'_{01}H'_{20} - H'_{10}H'_{11}) / (H'_{10}{}^2H'_{02} + H'_{01}{}^2H'_{20} - 2H'_{10}H'_{01}H'_{11}),$$

and that with these values of α and β

$$\text{var}(\eta^*/\sigma^2) = \frac{H'_{20}H'_{02} - H'_{11}{}^2}{H'_{10}{}^2H'_{02} + H'_{01}{}^2H'_{20} - 2H'_{10}H'_{01}H'_{11}} - 1.$$

Proceeding as indicated in §2, we find expansions of the form

$$\alpha = \sum_{i=1}^4 A_i(p_r)(n+1)^{-i} + O(n+1)^{-5},$$

$$\beta = \sum_{i=2}^5 B_i(p_r)(n+1)^{-i} + O(n+1)^{-6},$$

$$\text{var}(\eta^*/\sigma^2) = \sum_{i=1}^4 W_i(p_r)(n+1)^{-i} + O(n+1)^{-5}.$$

Values of $A_i(p_r)$, $B_i(p_r)$ and $W_i(p_r)$ are given in Table 3.

By comparing $W_1(p_r)(n+1)^{-1}$ with the large sample form of the variance of $\hat{\sigma}^2$, the maximum likelihood estimator of σ^2 , it is found that η^* has asymptotic efficiency of 100 % for all values of the ratio $r/(n+1)$.

It has not been found possible to give simple relations connecting $\alpha(r, n)$, $\beta(r, n)$, $\text{var} \eta^* |_{(r, n)}$ and $\alpha(n-r+1, n)$, $\beta(n-r+1, n)$, $\text{var} \eta^* |_{(n-r+1, n)}$ without bringing in auxiliary tables. However, it is hoped that the existing range on p_r will meet most requirements. In any case, an estimate of variance will be most unreliable when more than half the observations have been censored.

6. INTERPOLATION IN TABLES 2 AND 3

Owing to the labour involved in producing the tables, it was not found possible to give a fine entry for p_r . However, the following method of interpolation gives satisfactory results.

Suppose that n and r are given and that $p_r = r/(n+1)$ is not a tabulated value. Let $U(0.65)$ denote, for example, $\sum_{i=1}^5 E_i(0.65)(n+1)^{-i}$; then if $h = 20r/(n+1) - 13$, we may use as an approximation to $U(p_r)$ the expression

$$\frac{h(h^2-1)(h^2-4)(h^2-9)}{720} \left[\frac{U(0.50)}{h+3} - \frac{6U(0.55)}{h+2} + \frac{15U(0.60)}{h+1} - \frac{20U(0.65)}{h} \right. \\ \left. + \frac{15U(0.70)}{h-1} - \frac{6U(0.75)}{h-2} + \frac{U(0.80)}{h-3} \right].$$

This corresponds to Lagrange 7-point interpolation and is reasonably quick to apply.

As an example, using this method with $r = 7$, $n = 10$, we obtain

$$\left. \begin{aligned} \epsilon &= 0.38695262, \text{ whereas the true value from Table 1 is } 0.38695268; \\ \text{var}(\mu^*/\sigma) &= 0.11796547, \text{ whereas the true value from Table 1 is } 0.11796680; \\ \alpha &= 0.1938 \\ -\beta &= 0.01767 \\ \text{var}(\eta^*/\sigma^2) &= 0.3815 \end{aligned} \right\} \begin{array}{l} \text{No check is available on these values but they are thought to be} \\ \text{correct to the number of places of decimals given.} \end{array}$$

The formula was also used for extrapolation and applied to the power series for ϵ in the cases $n = 19$; $r = 17, 18$ and 19 . Checking against the true values in Table 1, the errors were found to be 1.10^{-5} , 3.10^{-5} and 2.10^{-4} , respectively.

7. A WORKED EXAMPLE

The following example is taken from Gupta's paper.

The table gives the day on which the first seven of ten tested mice died after being inoculated with a uniform culture of human tuberculosis.

Day	41	44	46	54	55	58	60
Log day = (x_i)	1.613	1.644	1.663	1.732	1.740	1.763	1.778
Frequency	1	1	1	1	1	1	1

It is found that

$$\bar{x}_6 = 1.6925, \quad x_7 = 1.778, \quad \sum_{i=1}^6 (x_i - x_7)^2 = 0.062191, \quad \left[\sum_{i=1}^6 (x_i - x_7) \right]^2 = 0.263169.$$

Thus

$$\mu^* = (0.3870)(1.6925) + (0.6130)(1.778) = 1.745,$$

$$\eta^* = (0.1938)(0.062191) - (0.01767)(0.263169) = 0.00740.$$

An estimate of the variance of μ^* is $(0.1180)\eta^* = 0.873 \times 10^{-3}$.

An estimate of the variance of η^* is $(0.3815)\eta^{*2} = 0.209 \times 10^{-4}$.

In passing, it is perhaps worth mentioning that μ^* has been found to be distributed very close to normality. Thus for rough tests on the hypothesis $\mu = \mu_0$, it may be possible to refer $(\mu^* - \mu_0)/\sqrt{(\text{est. var } \mu^*)}$ to the standard normal distribution. It seems intuitively more likely, however, that the distribution of this ratio will be better represented by the Pearson

Type IV curve. Similarly, for a rough test of the hypothesis $\sigma = \sigma_0$, we could refer η^* to the Pearson Type III distribution with the appropriate mean and standard deviation. These two problems have yet to be fully investigated and will probably form the subject of a later paper.

I wish to express my gratitude to Prof. E. S. Pearson, whose suggestions are responsible for several improvements on the original paper.

REFERENCES

- GUPTA, A. K. (1952). *Biometrika*, **39**, 260-73.
SARHAN, A. E. & GREENBERG, B. G. (1956). *Ann. Math. Statist.* **27**, 427-51.
SAW, J. G. (1958). *Biometrika*, **45**, 211-21. Corrigenda pp. 587-8.

Table 1. Values of the weight ϵ and the variance of μ^*/σ for $1 < r < n \leq 20$

	Weight, ϵ	var μ^*/σ		Weight ϵ	var μ^*/σ
$n = 3, r = 2$	Zero	0.4486 7110	$n = 11, r = 2$	-2.0245 4630	1.2178 6597
			3	-1.2238 3557	0.4573 7320
$n = 4, r = 2$	-0.4055 5159	0.5129 9280	4	-0.6960 0994	.2569 4929
3	.3093 2088	.2882 4812	5	-.3060 1488	.1763 0723
			6	Zero	0.1371 6243
$n = 5, r = 2$	-0.7411 0682	0.6111 2918	7	0.2492 5991	.1160 7211
3	Zero	.2868 3366	8	.4572 0905	.1040 6917
4	.4724 9056	.2185 7232	9	.6332 2913	.0970 3950
			10	.7830 2152	.0929 9029
$n = 6, r = 2$	-1.0260 6712	0.7186 4005	$n = 12, r = 2$	-2.1728 1803	1.3044 1143
3	-0.2676 7985	.3041 5115	3	-1.3678 0736	0.4859 3821
4	.2226 9148	.2089 5708	4	-0.8356 5351	.2720 5514
5	.5735 0999	.1774 4443	5	-.4420 2130	.1834 7202
			6	-0.1324 0995	0.1398 0858
$n = 7, r = 2$	-1.2733 1715	0.8263 6223	7	.1205 7511	.1158 5477
3	-0.5023 8193	.3302 2028	8	.3325 5203	.1018 4129
4	Zero	.2104 4686	9	.5131 4930	.0933 4070
5	.3642 6439	.1674 8290	10	.6685 3886	.0881 2541
6	0.6422 3186	0.1498 1238	11	0.8324 6899	0.0847 2881
			$n = 13, r = 2$	-2.3100 7581	1.3868 1251
$n = 8, r = 2$	-1.4915 3218	0.9309 9884	3	-1.5009 4676	0.5208 4280
3	-0.7109 1496	.3604 2823	4	-0.9654 0881	.2874 7620
4	-.1997 0421	.2180 7919	5	-.5686 7675	.1914 0681
5	.1737 4521	.1646 3021	6	-0.2560 4852	0.1434 7564
6	0.4623 9456	0.1409 3561	7	Zero	.1167 9900
7	.6920 0548	.1298 1932	8	.2152 0220	.1009 0906
			9	.3993 2369	.0910 3555
$n = 9, r = 2$	-1.6867 5767	1.0312 9527	10	.5587 4621	.0847 5964
3	-0.8983 5837	0.3924 5050	11	0.6977 1536	0.0807 6006
4	-.3802 8184	.2292 5388	12	.8188 8582	.0782 8340
5	Zero	.1661 0128	$n = 14, r = 2$	-2.4378 4620	1.4653 7120
6	0.2960 5268	0.1368 8836	3	-1.6250 7121	0.5513 4148
7	.5344 6005	.1221 7529	4	-1.0865 6888	.3029 9021
8	.7297 0989	.1146 2569	5	-0.6871 5159	.1998 4054
			6	-0.3719 4277	0.1478 6561
$n = 10, r = 2$	-1.8633 5148	1.1269 0294	7	-.1133 1319	.1186 0682
3	-1.0685 0751	0.4250 4749	8	.1045 5472	.1009 3894
4	-0.5448 8270	.2425 0203	9	.2915 1408	.0897 6990
5	-.1592 4349	.1703 1365	10	.4540 6085	.0825 0407
6	0.1423 7402	0.1359 7144	11	0.5966 2366	0.0777 1198
7	.3869 5268	.1179 6680	12	.7222 2478	.0745 6212
8	.5896 3408	.1080 8017	13	.8327 4358	.0725 5809
9	.7592 5335	.1026 6399			

Table 1 (cont.)

Table 1 (cont.)						
	Weight, ϵ	var μ^*/σ		Weight, ϵ	var μ^*/σ	
$n = 15, r =$	2	-2.5573 5741	$n = 18, r =$	6	-0.7737 6245	0.1692 4581
	3	-1.5910 5633		7	- .5077 6661	.1308 0404
	4	-1.2001 9350		8	- .2825 0321	.1064 6866
	5	-0.7984 1909		9	- .0879 4308	.0910 3752
				10	.0825 9311	.0802 8344
	6	-0.4809 6831		11	0.2337 6240	0.0727 5085
	7	- .2201 2308		12	.3689 2384	.0673 7399
	8	Zero		13	.4905 5815	.0634 8909
	9	.1893 1561		14	.6005 1947	.0606 6704
	10	.3543 9077		15	.7001 8374	.0586 2267
	11	0.4997 5262	16	0.7905 1758	0.0571 6368	
	12	.6285 8591	17	.8720 4112	.0561 6310	
	13	.7431 0067				
	14	.8446 4790				
$n = 16, r =$	2	-2.6696 1409	$n = 19, r =$	2	-2.9705 2895	1.8100 4988
	3	-1.8506 3947		3	-2.1441 9951	0.6910 0073
	4	-1.3071 6059		4	-1.5949 1464	.3780 2346
	5	-0.9032 9419		5	-1.1859 4337	.2445 7027
	6	-0.5838 7075		6	-0.8617 6868	0.1750 4878
	7	- .3210 9497		7	- .5944 0913	.1345 4868
	8	- .0990 2945		8	- .3678 0590	.1091 3104
	9	.0922 8706		9	- .1719 0350	.0923 2730
	10	.2594 7023		10	Zero	.0807 9098
	11	0.4071 0746		11	0.1525 8385	0.0726 4730
	12	.5384 7038	12	.2892 3182	.0667 8006	
	13	.6559 1509	13	.4124 5536	.0624 9205	
	14	.7610 9259	14	.5241 5083	.0593 3084	
	15	.8549 8345	15	.6257 5867	.0569 9385	
	$n = 17, r =$	2	-2.7754 4828	$n = 20, r =$	2	-3.0608 6752
3		-1.9538 0346	3		-2.2324 6540	0.7164 7801
4		-1.4082 0482	4		-1.6815 6241	.3925 8856
5		-1.0024 6439	5		-1.2711 8372	.2535 1721
6		-0.6812 8676	6		-0.9457 1364	0.1809 1765
7		- .4168 0820	7		- .6771 1925	.1384 3893
8		- .1930 4540	8		- .4493 0972	.1116 5405
9		Zero	9		- .2522 0650	.0938 5532
10		.1689 7949	10		- .0790 8873	.0815 6600
11		0.3185 2094	11		0.0747 4052	0.0728 3429
12		.4519 4855	12	.2126 8352	.0664 9576	
13		.5716 9627	13	.3372 7331	.0618 2136	
14		.6795 4769	14	.4504 3378	.0583 3664	
15		.7767 5745	15	.5536 4389	.0557 2300	
16		0.8640 4026	16	0.6480 4129	0.0537 6121	
$n = 18, r =$		2	-2.8755 5681	17	.7344 8217	.0522 9780
	3	-2.0514 7109	18	.8135 5787	.0512 2505	
	4	-1.5039 4667	19	.8860 9203	.0504 6728	
	5	-1.0965 1402				

Table 2. Coefficients of $(n+1)^{-i}$ in the series approximation to (a) the weight factor ϵ and (b) the variance, of the estimator of location, μ^*

$$(a) \epsilon = \sum_{i=0}^5 E_i(p_r) (n+1)^{-i} + O(n+1)^{-6} \quad \text{with} \quad p_r = r/(n+1)$$

p_r	$E_0(p_r)$	$E_1(p_r)$	$E_2(p_r)$	$E_3(p_r)$	$E_4(p_r)$	$E_5(p_r)$
0.50	Zero	Zero	Zero	Zero	Zero	Zero
0.55	0.14865772	0.0783132	0.006281	-0.04107	-0.0201	0.077
0.60	.28235913	.1375787	.001435	- .07785	- .0248	.161
0.65	0.40340719	0.1818639	-0.012337	-0.11338	-0.0159	0.273
0.70	.51356217	.2138610	- .034313	- .15044	.0100	.448
0.75	.61418208	.2352568	- .065129	- .19253	.0646	.763
0.80	.70631072	.2468760	- .107286	- .24566	.1784	1.435

$$(b) \text{var} \left(\frac{\mu^*}{\sigma} \right) = \sum_{i=1}^5 V_i(p_r) (n+1)^{-i} + O(n+1)^{-6}$$

p_r	$V_1(p_r)$	$V_2(p_r)$	$V_3(p_r)$	$V_4(p_r)$	$V_5(p_r)$
0.50	1.5707963	0.896605	0.09873	-0.4265	-0.132
0.55	1.4169573	.716520	.07237	- .1662	.275
0.60	1.3034728	.633049	.17656	.2218	.970
0.65	1.2186755	0.605297	0.32705	0.5913	1.481
0.70	1.1547730	.610085	.48554	.9072	1.840
0.75	1.1064195	.634026	.63859	1.1676	2.090
0.80	1.0698709	.669408	.78204	1.4575	2.333

Table 3. Coefficients of $(n+1)^{-i}$ in the series approximation to the weights α and β used in estimating σ^2 and for the variance of this estimator

p_r	$A_1(p_r)$	$A_2(p_r)$	$A_3(p_r)$	$A_4(p_r)$
0.50	2.482906	5.61883	6.8137	16.886
0.55	2.213248	4.34260	4.3813	10.188
0.60	1.989519	3.40329	2.7795	6.215
0.65	1.800852	2.69428	1.7316	3.852
0.70	1.639499	2.14687	0.9989	2.226
0.75	1.499774	1.71526	0.4764	1.276
0.80	1.377385	1.36769	0.0451	0.515

p_r	$B_2(p_r)$	$B_3(p_r)$	$B_4(p_r)$	$B_5(p_r)$
0.50	-1.517094	-9.70572	-11.8257	-32.504
0.55	-1.603473	-7.85637	-8.9535	-21.737
0.60	-1.604646	-6.42606	-6.7706	-14.903
0.65	-1.560119	-5.30056	-5.1900	-10.557
0.70	-1.491833	-4.40321	-3.9500	-7.335
0.75	-1.412189	-3.67797	-2.9867	-5.184
0.80	-1.328224	-3.08317	-2.1429	-3.438

p_r	$W_1(p_r)$	$W_2(p_r)$	$W_3(p_r)$	$W_4(p_r)$
0.50	4.965812	9.05846	16.1363	32.610
0.55	4.426496	7.11698	11.2447	20.516
0.60	3.979038	5.66638	7.9504	13.174
0.65	3.601703	4.55496	5.6512	8.494
0.70	3.278998	3.68410	4.0015	5.446
0.75	2.999548	3.00114	2.8366	3.585
0.80	2.754769	2.41742	1.8451	1.902

$$\alpha = \sum_{i=1}^4 A_i(p_r) (n+1)^{-i} + O(n+1)^{-5}, \quad \beta = \sum_{i=2}^5 B_i(p_r) (n+1)^{-i} + O(n+1)^{-6},$$

$$\text{var}(\eta^*/\sigma^2) = \sum_{i=1}^4 W_i(p_r) (n+1)^{-i} + O(n+1)^{-5}, \quad \text{with } p_r = r/(n+1).$$

ON HOTELLING'S GENERALIZATION OF T^2 BY K. C. SREEDHARAN PILLAI[†] AND PABLO SAMSON, JR.[‡]*The Statistical Center, University of the Philippines*

1. INTRODUCTION

Let S_1 and S_2 denote the matrices consisting of the sum of squares and products of deviations from means for two independent samples of sizes n_1 and n_2 , respectively, taken from p -variate normal populations. Then T_0^2 (Hotelling, 1951) may be defined by

$$T_0^2/(n_2 - 1) = \text{trace } (S_2^{-1} S_1).$$

It is a generalization of his T^2 statistic. Since the trace of a matrix equals the sum of its characteristic roots, $T_0^2/(n_2 - 1)$ is the sum of the characteristic roots of the matrix $S_2^{-1} S_1$. Define $U^{(s)}$ as the sum of the characteristic roots of $S_2^{-1} S_1$, where s is the number of non-zero roots. Thus $s = p$ if $n_1, n_2 > p$. It may be noted that $0 \leq U^{(s)} < \infty$.

We give here expressions for the moments of $U^{(s)}$ for $s = 2, 3, \dots, 6$. Upper 5 and 1 % points of the distribution of $U^{(s)}$ have been computed for $s = 2, 3$ and 4 for various values of two supplementary parameters using the exact moment quotients of $U^{(s)}$. Comparisons of these values for $s = 2$ are made with those obtained from the exact distribution, and those from an approximation by Pillai (1954). For $s = 3$ and 4, comparisons are made only with the approximation. A numerical example is also given to illustrate the use of the criterion $U^{(s)}$.

2. THE MOMENTS OF $U^{(s)}$

By making use of a recurrence relation given by Pillai (1954, 1956) the first four central moments of $U^{(s)}$ for $s = 2, 3, 4, 5$ and 6 have been obtained. We shall write

$$m = \frac{1}{2}(n_1 - s - 2), \quad n = \frac{1}{2}(n_2 - s - 2). \quad (2.1)$$

The nature of the first three moments suggests the following general forms:

$$\mu_1^{(s)} = \frac{s\{m + \frac{1}{2}(s+1)\}}{n}, \quad (2.2)$$

$$\mu_2^{(s)} = \frac{s\{m + \frac{1}{2}(s+1)\} \{n + m + \frac{1}{2}(s+1)\} (2n + s)}{n^2(n-1)(2n+1)}, \quad (2.3)$$

$$\mu_3^{(s)} = \frac{2s\{m + \frac{1}{2}(s+1)\} \{n + m + \frac{1}{2}(s+1)\} (n + 2m + s + 1) (2n + 2s) (2n + s)}{n^3(n-1)(n-2)(2n+2)(2n+1)}. \quad (2.4)$$

Although the fourth raw moment of $U^{(s)}$ has also been generalized (Pillai & Mijares, 1959), the central moment has not yet been obtained because of the heavy algebraic computations

[†] United Nations Senior Advisor in Mathematical Statistics and Visiting Professor of Statistics, Statistical Center, University of the Philippines.

[‡] Lecturer in Statistics, Statistical Center, University of the Philippines; portions of this paper are based on the M.A.S. thesis of the latter author submitted to the Statistical Center, March 1958.

involved. The fourth central moments of $U^{(s)}$ for $s = 2, 3$ and 4 which are used for computation in the following sections are therefore given in their explicit forms:

$$\mu_4^{(2)} = \frac{A_{42}}{n^4(n-1)(n-2)(n-3)(2n+1)(2n-1)}, \quad (2.5)$$

where

$$A_{42} = 3(2m+3)(2n+2m+3)\{[4n^2+(2m+3)(2n+2m+3)(n+3)](n+1)(n-2) + 16n(n+2m+3)^2\},$$

$$\mu_4^{(3)} = \frac{A_{43}}{(n+1)n^4(n-1)(n-2)(n-3)(2n+1)(2n-1)}, \quad (2.6)$$

where

$$A_{43} = 9(m+2)(n+m+2)\{(n+1)(n-2)[-2n^2+(m+2)(n+m+2)\{12(n+4)^2 - 88(n+4)+125\}] + (n+2m+4)^2[8(n+4)^3 - 38(n+4)^2 - 24(n+4)+134]\},$$

$$\mu_4^{(4)} = \frac{A_{44}}{(n+1)n^4(n-1)(n-2)(n-3)(2n+3)(2n+1)(2n-1)}, \quad (2.7)$$

where

$$A_{44} = 24(2m+5)(2n+2m+5)\{(n+1)(n-2)[2n^2+(2m+5)(2n+2m+5)\{(n+5)^3 - 12(n+5)^2 + 44(n+5) - 49\}] + (n+2m+5)^2[2(n+5)^4 - 17(n+5)^3 + 20(n+5)^2 + 111(n+5) - 224]\}.$$

Hence,

$$\beta_1^{(s)} = \frac{4(n+2m+s+1)^2(2n+2s)^2(n-1)(2n+1)}{s\{m+\frac{1}{2}(s+1)\}\{n+m+\frac{1}{2}(s+1)\}(2n+s)(n-2)^2(2n+2)^2}, \quad (2.8)$$

$$\beta_2^{(2)} = \frac{(n-1)(2n+1)A_{42}}{(2m+3)^2(2n+2m+3)^2(n+1)^2(n-2)(n-3)(2n-1)}, \quad (2.9)$$

$$\beta_2^{(3)} = \frac{(n-1)(2n+1)A_{43}}{9(m+2)^2(n+m+2)^2(2n+3)^2(2n-1)(n+1)(n-2)(n-3)}, \quad (2.10)$$

$$\text{and } \beta_2^{(4)} = \frac{(n-1)(2n+1)A_{44}}{4(2m+5)^2(2n+2m+5)^2(n+2)^2(2n+3)(2n-1)(n+1)(n-2)(n-3)}. \quad (2.11)$$

It may be observed that the moments of $U^{(s)}$ can be obtained in a simple manner from those of $V^{(s)}$, another criterion which is defined as the sum of $\theta_i = \lambda_i/(1+\lambda_i)$. In the numerator of any moment of $V^{(s)}$, if one maintains the coefficient of the highest power of n , then substitute and put alternatively positive and negative signs for subsequent powers of n , then substitute for n the value $m+n+s+1$, one obtains the numerator of the corresponding moment of $U^{(s)}$. Similarly, the denominator of any moment of $U^{(s)}$ can be derived from the corresponding moment of $V^{(s)}$ by putting positive coefficients for terms involving n and negative coefficients for other terms in each linear factor and making the substitution of $m+n+s+1$ for n .

3. THE UPPER PERCENTAGE POINTS OF THE DISTRIBUTION OF $U^{(2)}$

The moment quotients, $\beta_1^{(2)}$ and $\beta_2^{(2)}$, have been computed using (2.8) and (2.9) for parameters $m = 1(0.5) 5(5) 30; 50$ and $n = 15(5) 50; 60, 80, 100$ (except for a few values of n and smaller values of m). The upper 5 and 1 % points are then evaluated from 'Tables of the percentage points of Pearson curves for given β_1 and β_2 expressed in standardized measure' (Pearson &

Hartley, 1954, Table 42) and are shown in Tables 4 and 5 printed at the end of this paper.

Hotelling (1931, 1951) has obtained the exact C.D.F. of $U^{(2)}$ in terms of incomplete β -functions and it is given below:

$$P = I_w(2m+2, 2n+3) - \frac{(2n+2m+4)! \left(\frac{1-w}{1+w}\right)^{n+1}}{2(2m+1)! (2n+2)!} B_{w^2}(m+1, n+2), \quad (3.1)$$

where $w = U^{(2)}/(2 + U^{(2)})$.

Pillai (1954, 1956) has suggested an approximation to the distribution of $U^{(s)}$ in the form

$$p(U^{(s)}) = \frac{[U^{(s)}]^{s(m+\frac{1}{2}(s+1))-1}}{B\{s[m+\frac{1}{2}(s+1)], sn+1\} s^{s(m+\frac{1}{2}(s+1))} \left(1 + \frac{U^{(s)}}{s}\right)^{s(n+m+\frac{1}{2}(s+1))+1}} \quad (3.2)$$

and the approximate C.D.F. for $s = 2$ can be shown to be

$$P = I_w(2m+3, 2n+1). \quad (3.3)$$

A comparison of percentage points from the moment quotients with those from the exact and approximate C.D.F.'s can be judged from Table 1. A study of these tables confirms Pillai's observation (1954) that the error in using his approximate distribution is quite negligible for larger values of n ; for example, for the upper 5 % points, the approximate distribution could be used when n is 30 or more; for the upper 1 % points, when n is a little larger, say around 40. The percentage points from the moment quotients differ from those of the exact C.D.F. only in the third significant digits for large values of n , and appear relatively closer to the exact percentage points than the approximate percentage points. These observations justify the use of percentage points from the moment quotients or the use of the percentage points from the approximate C.D.F. for larger values of n .

Table 1. Comparison of approximations to the upper percentage points of $U^{(2)}$

m	n	5 % points			1 % points		
		From exact C.D.F.	From approx. C.D.F.	From β_1, β_2	From exact C.D.F.	From approx. C.D.F.	From β_1, β_2
0	5	1.43	1.39	—	2.17	2.05	—
0	10	0.671	0.664	—	0.979	0.931	—
4	10	1.927	1.886	—	2.516	2.415	—
0	15	0.443	0.435	—	0.624	0.602	—
4	15	1.236	1.219	—	1.563	1.526	—
0	20	0.328	0.323	—	0.455	0.443	—
3	20	0.7690	0.7600	0.7683	0.9698	0.9499	0.9702
3.5	20	0.839	0.829	0.842	1.050	1.029	1.059
4	20	0.908	0.898	0.907	1.130	1.108	1.130
5	20	1.046	1.034	1.045	1.288	1.263	1.288
0	30	0.2155	0.2124	—	0.2960	0.2904	—
3	30	0.5020	0.4983	0.5022	0.6235	0.6165	0.6236

4. UPPER PERCENTAGE POINTS OF THE DISTRIBUTION OF $U^{(s)}$ FOR CASES $s = 3$ AND 4

The moment quotients, β_1 and β_2 have been computed for various values of the parameters m and n using (2.8), (2.10) and (2.11) for $s = 3$ and 4. The percentage points were then again evaluated from Table 42 of Pearson & Hartley (1954) and are shown in Tables 6-9.

Making use of Pillai's approximate c.d.f. given by expression (3.2), we have for $s = 3$,

$$p = I_w(3m+6, 3n+1), \quad \text{where } w = U^{(3)}/(3+U^{(3)}) \quad (4.1)$$

and for $s = 4$

$$p = I_w(4m+10, 4n+1), \quad \text{where } w = U^{(4)}/(4+U^{(4)}) \quad (4.2)$$

For comparison with the upper 5 and 1 % points from the moment quotients, percentage points were also evaluated for certain values of m and n from the approximate c.d.f. given in (4.1) and (4.2). The results are tabulated together with the corresponding percentage points from β_1 and β_2 in Tables 2 and 3.

Table 2. Comparison of approximations to the upper percentage points of $U^{(3)}$

n	5 % points				1 % points			
	m = 0		m = 5		m = 0		m = 5	
	From β_1, β_2	From approx. c.d.f.	From β_1, β_2	From approx. c.d.f.	From β_1, β_2	From approx. c.d.f.	From β_1, β_2	From approx. c.d.f.
15	—	—	2.120	2.076	—	—	2.560	2.462
20	—	—	1.557	1.533	—	—	1.852	1.798
25	—	—	1.229	1.213	—	—	1.449	1.415
30	0.3621	0.3575	1.015	1.005	0.4635	0.4524	1.190	1.168
35	0.3097	0.3055	0.8638	0.8571	0.3939	0.3867	1.009	0.9943
40	.2692	.2671	.7520	.7483	.3412	.3376	0.8759	.8672
45	.2387	.2373	.6660	.6636	.3028	.2995	.7736	.7680
50	.2144	.2133	.5977	.5957	.2714	.2686	.6927	.6886
60	0.1778	0.1774	0.4959	0.4936	0.2250	0.2227	0.5728	0.5693
80	.1331	.1327	.3688	.3687	.1674	.1662	.4238	.4240
100	.1062	.1060	.2949	.2947	.1333	.1328	.3384	.3382

A study of these tables shows that the percentage points from the two methods are close for larger values of n and hence the approximate distributions appear usable as follows:

	5 %	1 %
$U^{(3)}$ $U^{(4)}$	$n \geq 40$ $n \geq 50$	$n \geq 50$ $n \geq 60$

Table 3. Comparison of approximations to the upper percentage points of $U^{(4)}$

n	5 % points				1 % points			
	$m = 0$		$m = 5$		$m = 0$		$m = 5$	
	From β_1, β_2	From approx. C.D.F.	From β_1, β_2	From approx. C.D.F.	From β_1, β_2	From approx. C.D.F.	From β_1, β_2	From approx. C.D.F.
15	1.122	1.084	2.877	2.812	1.413	1.334	3.395	3.246
20	0.8272	0.808	2.113	2.085	1.028	0.9841	2.458	2.387
25	.6548	.6415	1.668	1.611	0.8069	.7800	1.925	1.881
30	.5423	.5342	1.378	1.373	.6644	.6478	1.581	1.550
35	0.4629	0.4572	1.174	1.172	0.5646	0.5530	1.341	1.329
40	.4033	.3994	1.023	1.014	.4911	.4819	1.165	1.156
45	.3574	.3544	0.9058	0.9035	.4343	.4264	1.029	1.021
50	.3209	.3184	.8130	.8095	.3893	.3821	0.9218	0.9134
60	0.2666	0.2651	0.6745	0.6738	0.3224	0.3180	0.7627	0.7598
80	.1990	.1984	.5030	.5039	.2399	.2375	.5670	.5674
100	.1588	.1583	.4010	.4018	.1909	.1895	.4512	.4518

5. AN EXAMPLE

The criterion $U^{(s)}$ is useful for testing different kinds of hypotheses in multivariate analysis, one being the comparison of two independent estimates of the same dispersion matrix. A second problem in which this criterion can be used is in testing the equality of p -dimensional mean vectors of l p -variate normal populations (which is mathematically identical with the general problem of multivariate analysis of variance of means). The values of m and n appropriate for this test are

$$m = \frac{1}{2}|l - p - 1| - \frac{1}{2}, \quad n = \frac{1}{2}(N - l - p - 1), \tag{5.1}$$

where N is the total of l sample sizes. The following data are taken from Rao (1952, p. 263) concerning measurements of 140 schoolboys of almost the same age belonging to six different schools in an Indian city. Three characters were measured, namely, (1) head length, (2) height and (3) weight. The problem is to test for the significance of differences between schools in the mean characters. Let S^* , S be the matrices consisting of the sums of squares and products 'between' and 'within' schools for the three characters. Then

$$S^* = \begin{pmatrix} 1 & 2 & 3 \\ 752.0 & 214.2 & 521.3 \\ . & 151.3 & 401.2 \\ . & . & 1612.7 \end{pmatrix} \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} \tag{5.2}$$

$$S = \begin{pmatrix} 1 & 2 & 3 \\ 12,809.3 & 1003.7 & 2671.2 \\ . & 1499.6 & 4123.6 \\ . & . & 21,009.6 \end{pmatrix} \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} \tag{5.3}$$

The value of $U^{(3)} = 0.2016$ is obtained as the trace of the matrix $S^{-1}S^*$, where S^{-1} denotes the reciprocal of the S matrix. Referring to Tables 6 and 7, with $m = 0.5$ and $n = 65$, we find that this value of $U^{(3)}$ is just significant at the 5% level, which falls at about 0.195, and is clearly not significant at the 1% level, 0.242. The corresponding percentage points from Pillai's approximate c.d.f. are 0.194 and 0.239, respectively. This agrees with the findings of Rao, who examined the data using the Λ criterion of Pearson & Wilks (1933). Foster (1957), however, finds that the *largest root* is significant only at the 15% level.

The fact that in this particular example different test functions based on the same roots give different results need not be regarded as surprising. Far too little is known at present about alternatives to the null hypothesis to which these different tests are most sensitive.

The authors wish to acknowledge the facilities offered by the Statistical Center, University of the Philippines, during the course of preparation of this paper.

REFERENCES

- FOSTER, F. G. (1957). *Biometrika*, **44**, 441.
- HOTELLING, H. (1931). *Ann. Math. Statist.* **2**, 360.
- HOTELLING, H. (1944). Air testing of sample bombsights. Prepared for the Applied Mathematics Panel, National Defense Research Committee, by the Statistical Research Group, Columbia University, *SRG Report No. 328, AMP Report No. 5, 1R*.
- HOTELLING, H. (1951). A generalized T test and measure of multivariate dispersion. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, **2**, 23.
- PEARSON, E. S. & HARTLEY, H. O. (1954). *Biometrika Tables for Statisticians*, **1**. Cambridge University Press for the Biometrika Trustees.
- PEARSON, E. S. & WILKS, S. S. (1933). *Biometrika*, **25**, 353.
- PILLAI, K. C. S. (1954). On some distribution problems in multivariate analysis. *Mimeographed Series No. 88*. Institute of Statistics, University of North Carolina.
- PILLAI, K. C. S. (1956). *Ann. Math. Statist.* **27**, 1106.
- PILLAI, K. C. S. & MIJARES, T. A. (1959). *Ann. Math. Statist.* **30**.
- RAO, C. R. (1952). *Advanced Statistical Methods in Biometric Research*. New York: John Wiley and Sons, Inc.

Table 4. Upper 5% points of the sum of two non-null roots, $U^{(2)}$

$m \backslash n$	15	20	25	30	35	40	45	50	60	80	100
1.0	—	—	—	—	—	—	—	—	—	0.116	0.092
1.5	—	—	—	—	0.310	0.270	0.239	0.215	0.179	.133	.106
2.0	—	—	0.496	0.410	.350	.305	.270	.242	.202	.150	.120
2.5	—	—	.551	.456	.389	.339	.300	.269	.224	.167	.133
3.0	—	0.768	.607	.502	.428	.373	.330	.296	.246	.183	.146
3.5	—	0.842	0.662	0.547	0.466	0.406	0.360	0.323	0.268	0.200	0.159
4.0	—	.907	.717	.592	.505	.439	.389	.349	.290	.216	.172
4.5	—	.976	.771	.637	.542	.472	.418	.375	.311	.232	.185
5.0	—	1.045	.824	.681	.580	.505	.447	.401	.332	.248	.197
10	—	1.718	1.352	1.115	0.947	0.823	0.728	0.652	0.541	0.402	0.366
15	3.263	2.381	1.870	1.539	1.306	1.134	1.003	.899	.744	.553	.440
20	4.171	3.039	2.383	1.958	1.661	1.442	1.275	1.142	.945	.701	.558
25	5.077	3.692	2.894	2.374	2.014	1.749	1.545	1.383	1.143	.848	.674
30	5.981	4.345	3.403	2.792	2.366	2.053	1.813	1.622	1.341	.994	.790
50	9.590	6.947	5.430	4.449	3.767	3.266	2.881	2.577	2.127	1.573	1.248

Table 5. Upper 1% points of the sum of two non-null roots, $U^{(2)}$

$m \backslash n$	15	20	25	30	35	40	45	50	60	80	100
1.0	—	—	—	—	—	—	—	—	—	0.148	0.118
1.5	—	—	—	—	0.408	—	—	—	—	.168	.134
2.0	—	—	0.632	0.520	.442	.384	.339	0.273	0.226	.187	.149
2.5	—	—	.697	.572	.485	.422	.372	.303	.251	.205	.163
3.0	—	0.970	.760	.624	.529	.458	.406	.334	.276	.224	.178
3.5	—	1.059	0.822	0.675	0.572	0.497	0.439	.364	.301	.242	0.192
4.0	—	1.130	.884	.725	.616	.534	.472	0.393	0.325	.259	.206
4.5	—	1.210	.945	.776	.658	.571	.504	.422	.349	.277	.220
5.0	—	1.288	1.005	.825	.700	.607	.536	.451	.373	.294	.234
10	—	2.057	1.602	1.310	1.108	0.960	0.846	.480	.396	.294	0.366
15	3.934	2.812	2.183	1.782	1.505	1.302	1.147	0.757	0.625	0.462	.495
20	4.974	3.551	2.758	2.243	1.895	1.640	1.443	1.024	.845	.624	.621
25	6.043	4.302	3.330	2.712	2.286	1.974	1.736	1.288	1.061	.783	.745
30	7.092	5.043	3.900	3.173	2.673	2.306	2.028	1.549	1.275	.940	.868
50	11.28	7.996	6.167	5.007	4.202	3.626	3.185	2.837	2.328	1.711	1.353

Table 6. Upper 5% points of the sum of three non-null roots, $U^{(3)}$

$m \backslash n$	15	20	25	30	35	40	45	50	60	80	100
0.0	—	—	—	0.362	0.309	0.269	0.239	0.214	0.178	0.133	0.106
0.5	—	0.658	0.520	.431	.368	.321	.284	.255	.212	.158	.126
1.0	—	.761	.603	.499	.425	.371	.329	.295	.245	.183	.146
1.5	1.172	.864	.683	.566	.482	.420	.372	.334	.278	.207	.165
2.0	1.310	.964	.763	.632	.538	.469	.416	.373	.310	.231	.184
2.5	1.446	1.065	0.843	0.696	0.594	0.517	0.458	0.411	0.341	0.254	0.203
3.0	1.581	1.164	.921	.761	.649	.565	.500	.449	.372	.278	.222
3.5	1.717	1.263	.998	.825	.703	.612	.542	.486	.403	.293	.240
4.0	1.852	1.362	1.076	.889	.757	.659	.583	.524	.434	.324	.258
4.5	1.986	1.459	1.153	.952	.810	.706	.625	.561	.465	.347	.277
5.0	2.120	1.557	1.229	1.015	.864	.752	.666	.598	.496	.369	.295
10	3.438	2.517	1.982	1.634	1.391	1.210	1.071	0.961	0.796	0.593	0.472
15	4.742	3.494	2.725	2.245	1.933	1.661	1.469	1.316	1.090	.811	.646
20	6.039	4.407	3.478	2.852	2.424	2.107	1.862	1.668	1.381	1.026	.817
25	6.933	5.080	4.003	3.303	2.811	2.445	2.163	1.940	1.607	1.196	.953
30	8.168	5.981	4.711	3.886	3.306	2.875	2.543	2.278	1.888	1.406	1.120
50	10.10	7.327	5.734	4.708	3.990	3.460	3.053	2.731	2.255	1.671	1.326

N.B. For $m = -0.5$, $n = 100$, the 5% point of $U^{(3)}$ is 0.085.

Table 7. Upper 1% points of the sum of three non-null roots, $U^{(3)}$

$m \backslash n$	15	20	25	30	35	40	45	50	60	80	100
0.0	—	—	—	0.464	0.394	0.341	0.303	0.271	0.225	0.167	0.110
0.5	—	0.837	0.658	.541	.459	.400	.353	.317	.262	.195	.156
1.0	—	.956	.750	.617	.524	.456	.403	.361	.299	.223	.177
1.5	1.475	1.071	.840	.691	.587	.511	.451	.404	.335	.249	.198
2.0	1.634	1.186	.930	.765	.649	.564	.499	.447	.370	.275	.219
2.5	1.791	1.298	1.018	0.837	0.711	0.617	0.546	0.489	0.405	0.301	0.240
3.0	1.947	1.410	1.106	.909	.771	.670	.592	.531	.439	.326	.260
3.5	2.102	1.521	1.193	.980	.831	.722	.638	.572	.473	.339	.279
4.0	2.252	1.632	1.278	1.050	.891	.774	.684	.612	.506	.376	.299
4.5	2.408	1.742	1.364	1.120	.950	.825	.729	.653	.540	.401	.319
5.0	2.560	1.852	1.449	1.190	1.009	.876	.774	.693	.573	.424	.338
10	4.060	2.925	2.283	1.871	1.583	1.372	1.211	1.084	0.895	0.664	0.528
15	5.541	4.026	3.101	2.536	2.180	1.858	1.639	1.465	1.210	.896	.712
20	7.009	5.030	3.938	3.196	2.701	2.338	2.061	1.842	1.520	1.124	.893
25	7.838	5.658	4.419	3.618	3.066	2.658	2.345	2.098	1.732	1.285	1.021
30	9.212	6.639	5.184	4.244	3.593	3.114	2.747	2.455	2.029	1.505	1.196
50	11.69	8.318	6.432	5.237	4.406	3.803	3.344	2.982	2.451	1.805	1.428

Table 8. Upper 5% points of the sum of four non-null roots, $U^{(4)}$

$m \backslash n$	10	15	20	25	30	35	40	45	50	60	80	100
-0.5	—	—	0.712	0.547	0.453	0.387	0.337	0.299	0.268	0.223	0.167	0.133
0.0	—	1.122	.827	.655	.542	.463	.403	.357	.321	.267	.199	.159
0.5	—	1.302	.961	.761	.630	.537	.468	.415	.372	.309	.231	.184
1.0	—	1.482	1.092	.865	.716	.610	.531	.471	.426	.351	.262	.209
1.5	—	1.660	1.223	.968	.800	.682	.594	.526	.472	.392	.293	.234
2.0	—	1.837	1.352	1.120	.884	.753	.656	.581	.522	.433	.324	.258
2.5	—	2.012	1.480	1.171	0.967	0.824	0.718	0.636	0.571	0.474	0.354	0.282
3.0	—	2.186	1.608	1.271	1.050	.895	.780	.691	.620	.515	.384	.306
3.5	—	2.359	1.735	1.371	1.132	.963	.841	.745	.669	.555	.414	.330
4.0	—	2.532	1.862	1.470	1.214	1.035	.902	.799	.717	.595	.444	.354
4.5	—	2.710	1.988	1.569	1.296	1.104	.962	.852	.765	.636	.473	.378
5.0	—	2.877	2.113	1.668	1.378	1.174	1.023	.906	.813	.674	.503	.401
10	7.17	4.578	3.353	2.644	2.183	1.855	1.618	1.430	1.284	1.064	0.793	0.632
15	9.84	6.266	4.581	3.610	2.978	2.533	2.203	1.949	1.748	1.448	1.078	.859
20	12.15	7.404	5.803	4.570	3.768	3.204	2.784	2.463	2.207	1.828	1.361	1.083
25	15.15	9.622	7.021	5.526	4.553	3.869	3.363	2.974	2.665	2.207	1.641	1.306
30	17.80	11.30	8.238	6.481	5.337	4.533	4.050	3.464	3.121	2.583	1.920	1.528
50	28.40	17.98	13.09	10.29	8.464	7.184	6.238	5.505	4.938	4.082	3.031	2.410

Table 9. Upper 1% points of the sum of four non-null roots, $U^{(4)}$

$m \backslash n$	10	15	20	25	30	35	40	45	50	60	80	100
-0.5	—	—	0.908	0.687	0.565	0.480	0.418	0.370	0.331	0.274	0.204	0.163
0.0	—	1.413	1.028	.807	.664	.565	.491	.434	.389	.322	.240	.191
0.5	—	1.620	1.177	.925	.761	.647	.562	.497	.446	.369	.275	.219
1.0	—	1.824	1.325	1.040	.856	.727	.632	.559	.501	.415	.309	.246
1.5	—	2.025	1.470	1.154	.949	.807	.701	.620	.556	.460	.342	.272
2.0	—	2.224	1.615	1.322	1.042	.885	.769	.680	.609	.504	.375	.299
2.5	—	2.421	1.758	1.378	1.134	0.963	0.836	0.739	0.662	0.548	0.408	0.325
3.0	—	2.618	1.899	1.489	1.225	1.039	.902	.798	.715	.592	.440	.350
3.5	—	2.813	2.039	1.599	1.314	1.116	.969	.856	.767	.635	.472	.376
4.0	—	3.008	2.180	1.708	1.402	1.191	1.034	.914	.819	.678	.504	.401
4.5	—	3.202	2.319	1.817	1.493	1.266	1.100	.972	.870	.722	.535	.426
5.0	—	3.395	2.458	1.925	1.581	1.341	1.165	1.029	.922	.763	.567	.451
10	8.57	5.305	3.827	2.991	2.452	2.075	1.803	1.591	1.425	1.179	0.875	0.696
15	11.66	7.189	5.177	4.038	3.309	2.802	2.430	2.144	1.919	1.585	1.175	.934
20	14.74	9.065	6.520	5.079	4.159	3.521	3.050	2.690	2.406	1.986	1.472	1.169
25	17.81	10.94	7.857	6.115	5.004	4.233	3.666	3.233	2.890	2.385	1.767	1.403
30	20.88	12.80	9.19	7.15	5.85	4.94	4.40	3.74	3.37	2.78	2.06	1.63
50	33.13	20.26	14.51	11.27	9.20	7.77	6.72	5.91	5.29	4.36	3.22	2.55

TABLES FOR WALD TESTS FOR THE MEAN OF A NORMAL DISTRIBUTION

BY ELIZABETH D. BARRACLOUGH AND E. S. PAGE

Durham University Computing Laboratory

1. INTRODUCTION

If x is a normally distributed random variable with unknown mean θ and known standard deviation σ , the Wald sequential test for the hypothesis $\theta = \theta_0$ against the alternative $\theta = \theta_1$ ($\theta_1 > \theta_0$) is specified as follows (Wald, 1947, pp. 117 *et seq.*):

Continue taking observations x_i as long as

$$a < \frac{\theta_1 - \theta_0}{\sigma^2} \sum_{i=1}^n x_i + \frac{n}{2\sigma^2} (\theta_0^2 - \theta_1^2) < b; \quad (1)$$

if the left-hand inequality first fails to be satisfied, stop sampling and accept $\theta = \theta_0$, similarly accept $\theta = \theta_1$ if the right-hand inequality fails.

In practice the test is usually carried out either by plotting the cumulative sum $\sum x_i$ on a chart on which the parallel straight line boundaries are drawn and stopping sampling when one of the boundaries has been crossed or by comparing the cumulative sum after each observation with acceptance and rejection numbers calculated from (1).

$$\text{If } y_i = \{2x_i - (\theta_0 + \theta_1)\}/2\sigma, \quad (2)$$

$$\text{the inequalities (1) may be written } -Z < \sum_{i=1}^n y_i < h - Z, \quad (3)$$

$$\text{where } Z = -\frac{a\sigma}{\theta_1 - \theta_0}, \quad h = \frac{(b-a)\sigma}{\theta_1 - \theta_0}. \quad (4)$$

$$\text{The variate } y_i \text{ is normally distributed with unit standard deviation and mean } \mu = \{2\theta - (\theta_0 + \theta_1)\}/2\sigma, \quad (5)$$

where $E(x) = \theta$.

Let $P(Z | \mu)$ be the probability of accepting the hypothesis that $\theta = \theta_0$ when the true mean has a particular value θ , and μ is given by equation (5). The argument μ will be omitted where no ambiguity is possible. The derivation of the equation for $P(Z)$ given by Anscombe (1948) considers expectations conditional upon the first observation (Page, 1954*a*) and in this case yields

$$P(Z) = \Phi(-\mu - Z) + \int_0^h P(x) (2\pi)^{-\frac{1}{2}} \exp -\frac{1}{2}(x - Z - \mu)^2 dx, \quad (6)$$

where

$$\Phi(x) = \int_{-\infty}^x (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}t^2} dt.$$

If $\mu = \mp \frac{1}{2}(\theta_1 - \theta_0)/\sigma$, $P(Z | \mu)$ gives the probabilities of accepting $\theta = \theta_0$ when $E(x) = \theta_0$ and $E(x) = \theta_1$, respectively. The charts and Tables 1*a-d*, of the Appendix enable h and Z to be selected to give specified values for these two probabilities. The constants of the test, a and b , can then be calculated from equations (4). In Tables 1*a-d*,

$$P_- = P(Z | -\frac{1}{2}(\theta_1 - \theta_0)/\sigma) = 1 - \alpha,$$

$$P_+ = P(Z | \frac{1}{2}(\theta_1 - \theta_0)/\sigma) = \beta.$$

If the required values of the operating characteristic are not among those tabulated, it may be sufficient for practical purposes to interpolate by eye in the charts provided. The values in the tables were those requested for a particular application, but the range covered is sufficiently wide to be of more general use. Further combinations of α and β can be obtained by using the relationships

$$P(h - Z \mid -\frac{1}{2}(\theta_1 - \theta_0)/\sigma) = 1 - P(Z \mid \frac{1}{2}(\theta_1 - \theta_0)/\sigma),$$

$$P(h - Z \mid \frac{1}{2}(\theta_1 - \theta_0)/\sigma) = 1 - P(Z \mid -\frac{1}{2}(\theta_1 - \theta_0)/\sigma).$$

Tables 2a-d, of the Appendix, computed from equation (9) below, give the average sample numbers, N_- (when $\theta = \theta_0$) and N_+ (when $\theta = \theta_1$) of each one of these tests.

2. THE METHOD OF USE OF THE CHARTS AND TABLES

(a) To select a test with given risks

If a test of the mean of a normal distribution with standard deviation σ is required which has probability $1 - \alpha$ of accepting the hypothesis $\theta = \theta_0$ when $\theta = \theta_0$, and probability β when $\theta = \theta_1$ (so that the risks of the first and second kinds are α, β , respectively) it may be found as follows.

Calculate μ_0, μ_1 where
$$\mu_0 = -\mu_1 = (\theta_0 - \theta_1)/2\sigma. \quad (7)$$

Charts are given for $\mu = \pm 0.25, \pm 0.5, \pm 0.75, \pm 1.0$, and there are curves for $P(Z \mid \mu > 0) = 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70$ and $P(Z \mid \mu < 0) = 0.95, 0.99, 0.995, 0.999$.

Example 1

To select a test with $P(Z \mid \theta_0 = 3.0) = 0.99, P(Z \mid \theta_1 = 3.75) = 0.05$ when $\sigma = 0.5$.

We have

$$\mu_0 = -\mu_1 = (3.0 - 3.75)/2(0.5) = -0.75.$$

From Table 1c for $\mu = \pm 0.75, h = 3.88; Z = 1.42$.

Hence $a = -2.13, b = 3.69$ from (4). The required test is therefore to take observations as long as

$$-0.71 + 3.375n < \sum x_i < 1.23 + 3.375n.$$

(b) Operating characteristics

Several points on the operating characteristic (o.c.) may be obtained approximately by visual interpolation between the curves on the charts for the different values of μ .

Example 2

The approximate o.c. for the test selected in Example 1 is shown in Fig. 1 in which the circled points are those read from the charts.

(c) Average sample number

The average numbers of observations required for the tests given in Table 1 when the mean is that specified by either hypothesis are shown in Table 2.

Example 3

The average number of observations required by the test of Example 1 when $\theta = 3.0$ is $N = 2.87$ and when $\theta = 3.75$ is $N = 4.00$.

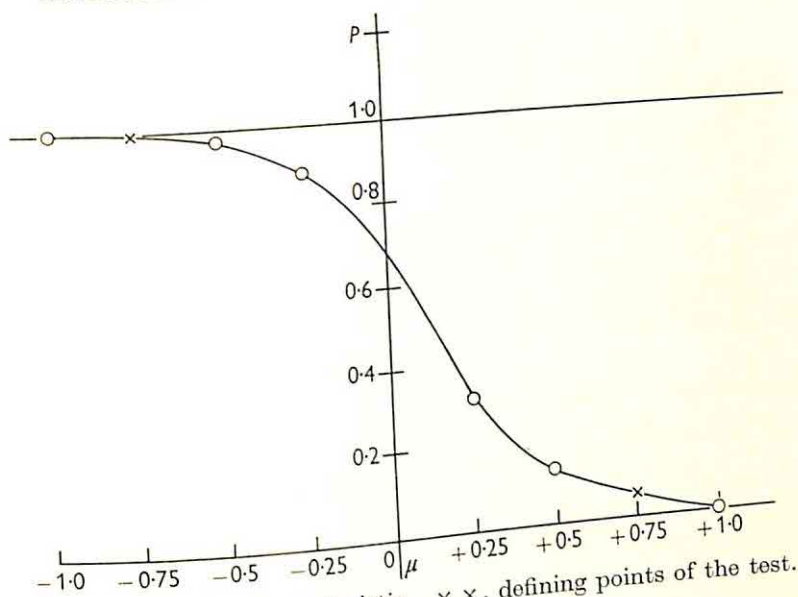


Fig. 1. Operating characteristic. $\times \times$, defining points of the test.
 $\circ \circ$, points read from the charts.

3. CALCULATION OF THE TABLES AND CHARTS

For a given pair of values of h and μ the solution $P(Z)$ to (6) was computed by an iterative method (Page, 1954*b*; Good, 1957). The known function, the normal distribution function, was calculated by a rational function approximation (Hastings, 1952) with an error at most $\frac{1}{2} \times 10^{-6}$. The integral containing the unknown solution was replaced by a Gaussian quadrature formula using the smallest number of ordinates that would give a fairly accurate approximation:

$$P(Z) = \Phi(-\mu - Z) + \sum_{i=1}^n w_i P(x_i) K(x_i, Z), \quad (8)$$

where w_i, x_i are respectively the Gauss weights and abscissae, and $K(x, Z)$ is the kernel of (6). Equation (8) was used iteratively to obtain $P(Z)$, for $Z = x_i, i = 1, \dots, n$, starting from an initial approximation (usually zero). The iteration continued until successive values of the $P(x_i)$ all differed by less than a small quantity. At this stage equation (8) was applied for $Z = y_j, j = 1, \dots, n'$, where y_j are the abscissae of a Gauss quadrature formula with more ordinates, $n' > n$. The $P(y_j)$ so obtained formed the starting value for iterations with the more accurate formula. This process continued automatically until the $P(Z)$ values at a set of Gauss points were reproduced to a prescribed accuracy by successive quadrature formulae. Values of $P(Z)$ at other points were calculated and used by inverse interpolation to obtain estimated Z values for the given $P(Z)$. These estimates were refined by substitution in (8) and repeating the process.

Some of the pivotal values of the $P(Z)$ were checked by an independent solution of (8) by the finite difference methods of Fox & Goodwin (1953). Limitations of storage space prevented attainment of the same accuracy of evaluation of the integral, but in no case did the corresponding pivotal values differ by more than one unit in the fourth decimal place from those derived by the first method.

The average sample number $N(Z)$ satisfies the equation

$$N(Z) = 1 + \int_0^h N(x) (2\pi)^{-\frac{1}{2}} \exp -\frac{1}{2}(x - Z - \mu)^2 dx, \quad (9)$$

which was solved by the first method for the tests quoted in Table 1.

4. WALD'S APPROXIMATIONS

Wald's approximation $P^*(Z)$ for the operating characteristic $P(Z)$ of the test may be written

$$P^*(Z) = (e^{-2\mu h} - e^{-2\mu Z}) / (e^{-2\mu h} - 1); \tag{10}$$

accordingly the test calculated from this approximation to give risks α, β , i.e. so that $P^*(Z | -\mu) = 1 - \alpha, \quad P^*(Z | \mu) = \beta$

has
$$Z = \frac{1}{2\mu} \log \frac{1 - \alpha}{\beta}, \quad h = \frac{1}{2\mu} \log \frac{(1 - \alpha)(1 - \beta)}{\alpha\beta}. \tag{11}$$

Wald also gives an approximation $N^*(Z)$ for the average sample number of the test.

$$N^*(Z | \mu) = [h\{1 - P^*(Z | \mu)\} - Z] / \mu. \tag{12}$$

Thus for the test above

$$N^*(Z | -\mu) = (Z - h\alpha) / \mu, \tag{13}$$

$$N^*(Z | \mu) = \{h(1 - \beta) - Z\} / \mu. \tag{14}$$

A comparison of $P^*(Z | \mu)$ and $P(Z | \mu)$ for one value of h , two values of Z and a range of μ was made by Page (1954*a*). The present table allows a more extensive comparison. Text-table 1 shows the Wald approximations α^*, β^* to the risks calculated from (10) for some tests with risks α, β at means $\pm \mu$. For $\mu = 0.25$ the approximate risks of the second kind, β , are about one-third greater than their correct values, while for $\mu = 1$ they are more than three times greater. Further, for $\mu = 1$, no approximations are possible when $\alpha = 0.001, \beta = 0.3$, for example, since $Z < 0$ and (10) produces a result greater than unity. A similar behaviour is shown for risks of the first kind. It is clear that for any work requiring an accurate assessment of the risks for $\mu \geq 0.25$ the Wald approximation cannot be accepted. Further, some of the Wald approximations exceed the true o.c. for $\mu < 0$ if Z is small; these are not likely to occur for the most frequently used schemes with small α and β .

Text-table 1

μ	h	Z	True o.c.	Wald o.c.	True A.S.N.	Wald A.S.N. with Wald o.c.	Wald A.S.N. with true o.c.
- 0.25	11.77	0.37	0.999	0.9994	4.32	1.4	1.4
- .25	14.90	5.40	.995	.992	23.84	21.1	21.3
- .25	10.61	5.31	.95	.934	21.46	18.4	19.1
+ 0.25	11.77	0.37	0.60	0.831	16.99	6.5	17.4
+ .25	14.90	5.40	.05	.067	37.43	34.0	35.0
+ .25	10.61	5.31	.05	.066	21.46	18.4	19.1
- 1.0	3.31	0.48	0.999	0.9979	1.49	0.5	0.5
- 1.0	1.72	0.86	.95	.848	1.63	0.6	0.8
+ 1.0	3.31	0.48	0.10	0.382	3.24	1.6	2.5
+ 1.0	1.72	0.86	.05	.152	1.63	0.6	0.8

Comparison of exact and approximate risks

The above comparison shows that tests selected by the Wald formulae to have specified risks will in fact usually have risks less than those specified by amounts depending largely upon the population mean. It would be useful to make an extensive comparison of the exact average sample numbers (A.S.N.) of the exact and approximate tests with specified risks, i.e. those with h, Z , chosen from (6) and (8), respectively. This would, however, entail more computation than is warranted. Instead we compare the approximate A.S.N. for the exact test with values h, Z taken from the tables and the approximate A.S.N. for the approximate test with values h, Z given by (9) with the true values of the A.S.N.

The approximate A.S.N. using the Wald approximation to the o.c. is always smaller than the true value for the exact test and the difference is frequently considerable. If, however, the true o.c. is used in (12) a better approximation to the true A.S.N. is obtained, but it is still too small in most cases and in particular no such approximation is useful if $Z < 0$. For $\mu = 0.25$, this approximation seems to yield an error of less than 10 %, but as μ increases the approximation deteriorates.

Improved approximations to the o.c. and A.S.N. have been derived by modifying the integral equations (6) and (9) (Page, 1954*a*) and more recently by a similar method by Kemp (1958). In the examples given, Kemp's approximations to $P(0), N(0)$ are usually nearer to the true values than Page's* but the latter are closer approximations for a range of Z near the middle of $(0, h)$. No further comparisons of these approximations have been made, nor of their quality over the wider range of schemes now available.

We wish to thank Dr O. L. Davies and the Pharmaceuticals Division of Imperial Chemical Industries for their financial support for the preparation of the graphs.

REFERENCES

- ANSCOMBE, F. J. (1948). Cambridge lecture notes.
 FOX, L. & GOODWIN, E. T. (1953). *Phil. Trans. A*, **245**, 501-34.
 GOOD, I. J. (1957). *Math. Tab., Wash.*, **11**, 82-3.
 HASTINGS, L. J. (1952). *Rational Approximations in Numerical Analysis*. Rand Corporation, Santa Monica. (Subsequently published by Princeton University Press, 1955.)
 KEMP, K. W. (1958). *J. Roy. Statist. Soc. B*, **20**, 379-86.
 PAGE, E. S. (1954*a*). *J. Roy. Statist. Soc. B*, **16**, 136-9.
 PAGE, E. S. (1954*b*). *Proc. Camb. Phil. Soc.* **50**, 414-25.
 WALD, A. (1947). *Sequential Analysis*. New York: John Wiley and Sons.
 WETHERILL, G. B. (1957). Ph.D. Thesis. University of London.

* Wetherill (1957) has found that the values of $N(0)$ for $k = 0.25, 0.50$ in Table 4 of Page (1954*a*) have been miscalculated and that the corrections give better approximations to the true values.

APPENDIX

Table 1

$P_- \backslash P_+$	0.999		0.995		0.99		0.95	
	h	Z	h	Z	h	Z	h	Z
(a) $\mu = 0.25$								
0.70	10.80	-0.03	7.57	-0.04	6.17	-0.05	2.81	-0.19
.60	11.77	0.37	8.55	0.36	7.15	0.35	3.82	0.24
.50	12.63	0.78	9.40	0.78	8.01	0.76	4.70	0.67
.40	13.45	1.25	10.23	1.24	8.84	1.23	5.53	1.15
0.30	14.22	1.83	11.12	1.82	9.72	1.81	6.42	1.73
.20	14.98	2.64	12.19	2.63	10.80	2.62	7.50	2.53
.10	—	—	13.78	4.01	12.42	4.00	9.12	3.92
.05	—	—	14.90	5.40	13.87	5.39	10.61	5.31
(b) $\mu = 0.5$								
0.70	4.42	-0.71	2.79	-0.72	2.08	-0.74	0.30	-0.95
.60	5.02	-0.39	3.40	-0.40	2.69	-0.42	0.97	-0.54
.50	5.54	-0.09	3.93	-0.10	3.22	-0.11	1.54	-0.19
.40	6.04	0.22	4.42	0.22	3.72	0.21	2.05	0.15
0.30	6.55	0.57	4.93	0.57	4.23	0.56	2.57	0.51
.20	7.13	1.02	5.51	1.02	4.81	1.01	3.16	0.97
.10	7.95	1.73	6.34	1.72	5.64	1.72	3.99	1.68
.05	8.69	2.42	7.08	2.41	6.38	2.41	4.73	2.37
(c) $\mu = 0.75$								
0.70	2.11	-1.12	1.01	-1.15	0.56	-1.18	—	—
.60	2.59	-0.82	1.50	-0.84	1.03	-0.86	—	—
.50	3.02	-0.55	1.93	-0.56	1.46	-0.57	0.27	-0.69
.40	3.42	-0.27	2.34	-0.28	1.87	-0.29	0.71	-0.37
0.30	3.83	0.03	2.75	0.03	2.28	0.02	1.14	-0.04
.20	4.28	0.40	3.20	0.40	2.74	0.39	1.62	0.35
.10	4.90	0.94	3.82	0.94	3.36	0.93	2.25	0.90
.05	5.42	1.42	4.34	1.42	3.88	1.42	2.78	1.39
(d) $\mu = 1.0$								
0.70	0.85	-1.45	—	—	—	—	—	—
.60	1.26	-1.16	0.47	-1.20	0.11	-1.24	—	—
.50	1.64	-0.90	0.83	-0.92	0.47	-0.94	—	—
.40	2.00	-0.63	1.19	-0.64	0.83	-0.66	—	—
0.30	2.36	-0.34	1.55	-0.35	1.20	-0.36	0.25	-0.44
.20	2.77	-0.00	1.97	-0.01	1.61	-0.01	0.69	-0.07
.10	3.31	0.48	2.51	0.48	2.16	0.47	1.27	0.44
.05	3.75	0.89	2.95	0.89	2.60	0.89	1.72	0.86

Table 2

$P_- \backslash P_+$	0.999		0.995		0.99		0.95	
	N_-	N_+	N_-	N_+	N_-	N_+	N_-	N_+
(a) $\mu = 0.25$								
0.70	3.15	12.00	3.00	8.16	2.84	6.53	2.00	2.92
.60	4.32	16.99	4.15	11.86	3.97	9.67	2.97	4.72
.50	5.75	22.34	5.56	15.92	5.37	13.17	4.23	6.90
.40	7.53	28.00	7.32	20.31	7.11	17.00	5.84	9.41
0.30	9.82	33.72	9.61	25.05	9.38	21.18	7.97	12.27
.20	13.06	39.10	12.83	30.17	12.58	25.74	11.00	15.51
.10	—	—	18.34	35.69	18.06	30.82	16.22	19.27
.05	—	—	23.84	37.43	23.54	33.51	21.46	21.46
(b) $\mu = 0.5$								
0.70	1.29	3.35	1.25	2.39	1.21	1.99	1.04	1.11
.60	1.49	4.53	1.45	3.25	1.41	2.71	1.18	1.49
.50	1.77	5.83	1.73	4.22	1.68	3.53	1.41	1.97
.40	2.16	7.22	2.11	5.29	2.05	4.47	1.74	2.57
0.30	2.69	8.71	2.63	6.47	2.58	5.50	2.22	3.27
.20	3.48	10.32	3.42	7.74	3.35	6.64	2.96	4.08
.10	4.86	12.06	4.80	9.16	4.72	7.91	4.26	5.02
.05	6.25	13.00	6.18	9.95	6.10	8.63	5.58	5.58
(c) $\mu = 0.75$								
0.70	1.05	1.80	1.04	1.38	1.02	1.21	—	—
.60	1.10	2.28	1.08	1.71	1.07	1.47	—	—
.50	1.18	2.81	1.16	2.10	1.14	1.79	1.03	1.12
.40	1.29	3.40	1.27	2.55	1.25	2.18	1.12	1.35
0.30	1.47	4.05	1.44	3.05	1.42	2.62	1.27	1.63
.20	1.76	4.75	1.73	3.61	1.71	3.11	1.53	1.97
.10	2.33	5.52	2.30	4.23	2.27	3.68	2.06	2.39
.05	2.94	5.94	2.91	4.59	2.87	4.00	2.64	2.64
(d) $\mu = 1.0$								
0.70	1.01	1.29	—	—	—	1.04	—	—
.60	1.02	1.52	1.01	1.19	1.00	1.21	—	—
.50	1.04	1.79	1.03	1.38	1.02	1.40	—	—
.40	1.07	2.09	1.06	1.61	1.05	1.63	1.03	1.09
0.30	1.13	2.44	1.12	1.87	1.11	1.89	1.12	1.27
.20	1.24	2.82	1.23	2.17	1.21	2.20	1.35	1.49
.10	1.49	3.24	1.48	2.52	1.46	2.38	1.63	1.63
.05	1.80	3.48	1.78	2.71	1.76	—	—	—

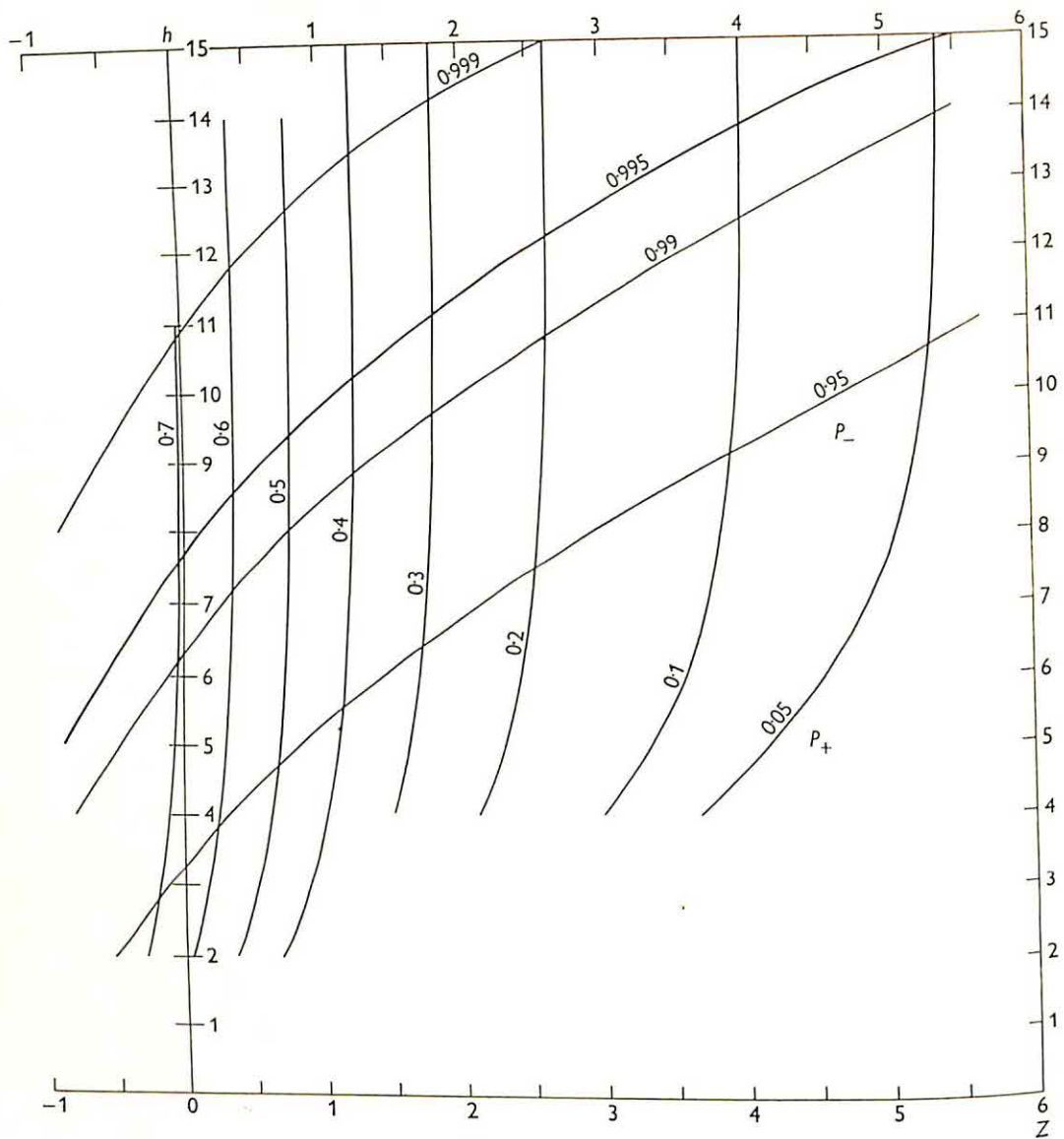


Chart I, $\mu = 0.25$

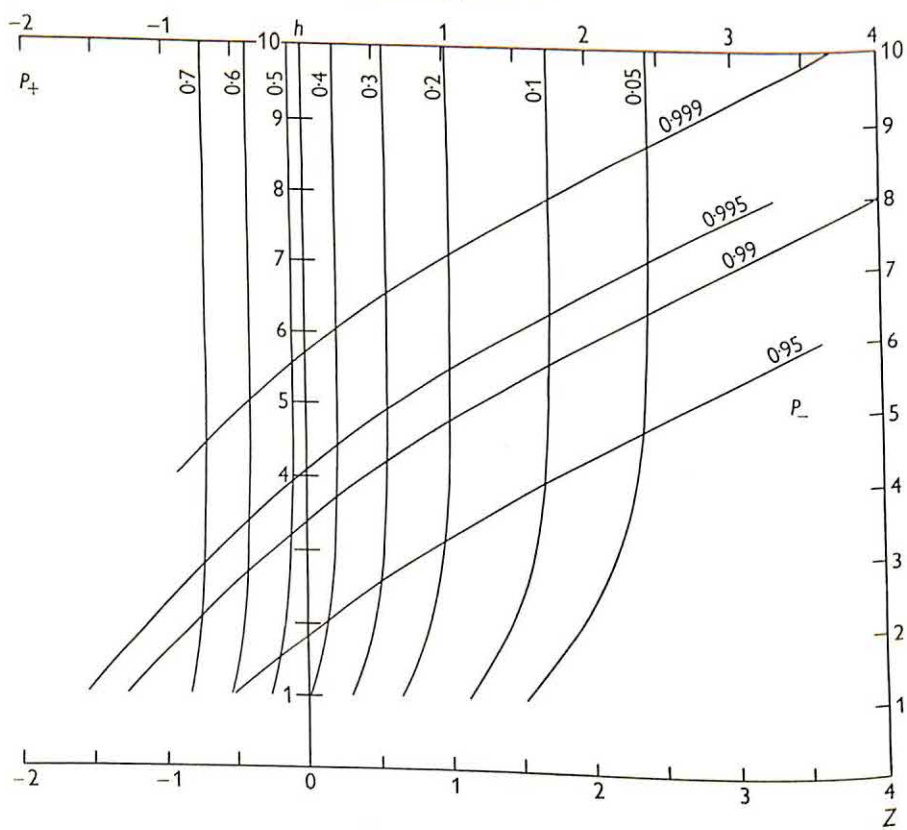


Chart II, $\mu = 0.5$

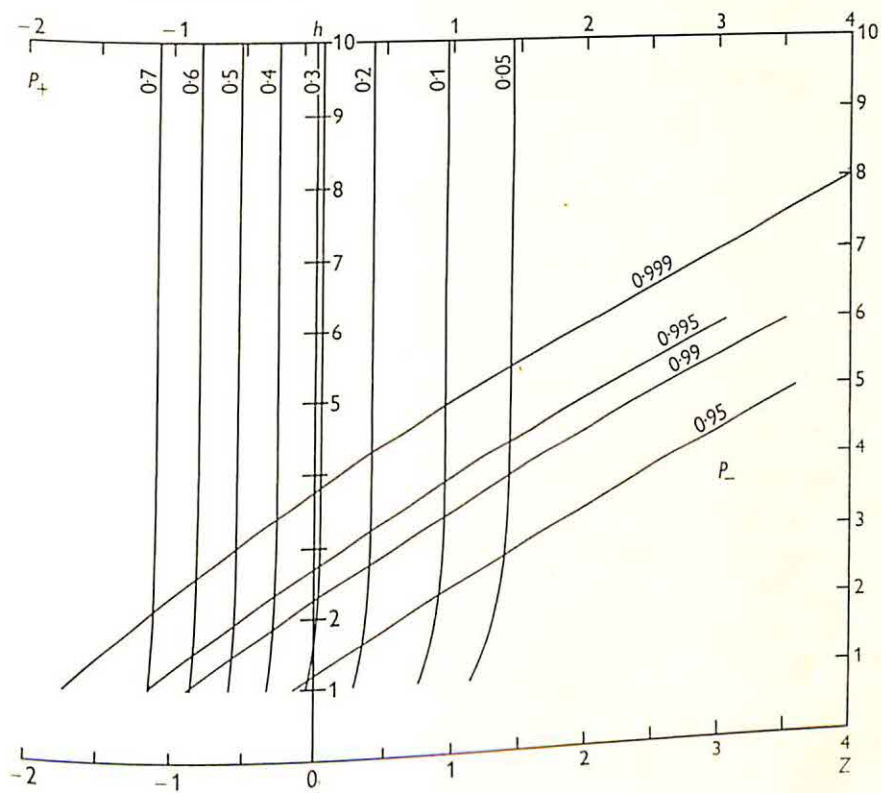


Chart III, $\mu = 0.75$

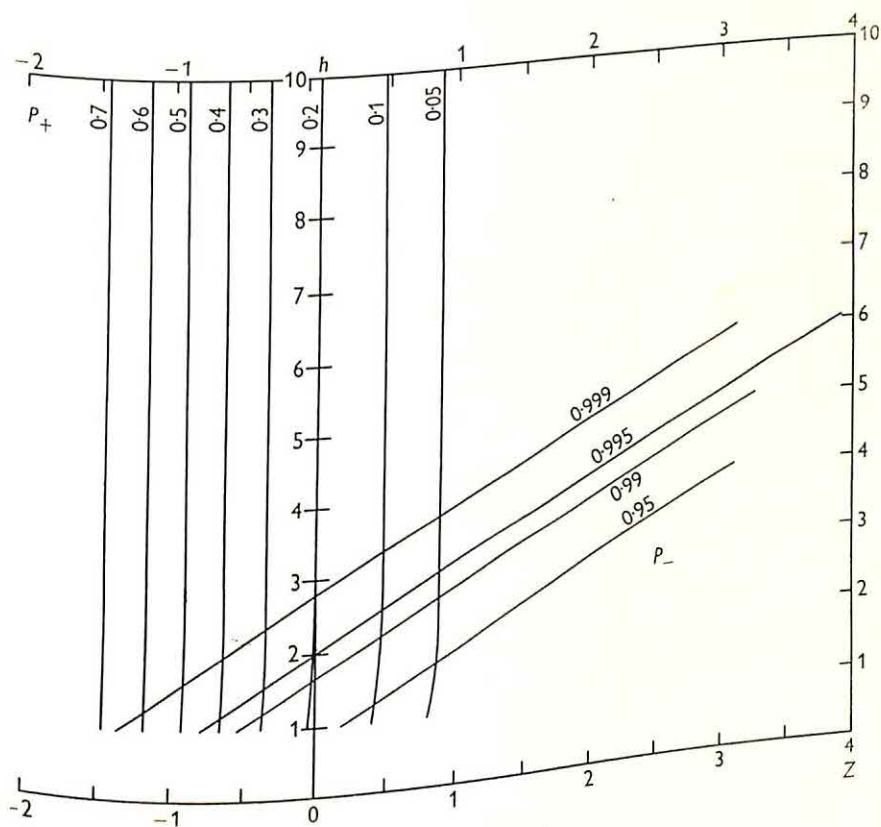


Chart IV, $\mu = 1.0$

TABLES OF RANDOM OBSERVATIONS FROM STANDARD DISTRIBUTIONS

BY M. H. QUENOUILLE

Research Techniques Unit, London School of Economics

PURPOSES OF THE TABLES

The recent increase in the popularity of Monte Carlo methods has been accompanied by a corresponding increase in the supply of random observations. These are now available not only in the original form of random numbers (Kendall & Babington Smith, 1939; Tippett, 1927; Rand Corp. 1955) but also as random normal deviates (Wold, 1954), correlated random normal deviates (Fieller, Lewis & Pearson, 1955) and serially correlated random numbers and normal deviates (Kendall, 1949).

From these published tables, it is easily possible to draw random observations from any distribution, the only steps involved being the calculation of the distribution function and the transformation of rectangularly distributed observations using this function. Nevertheless, these two steps may require considerable calculations, and it is therefore convenient to have readily available sets of random observations from some standard distributions. These are provided in this table.

CONTENTS OF THE TABLES

For Monte Carlo methods where the purpose is to investigate the sampling effects of departures from the normal (or any other standard) distribution, it is likely in most instances that the greatest information will be obtained by relating estimated values for the non-normal observations to the corresponding values obtained with the *same* observations transformed to normality. The table has therefore been arranged so that the values in each line provide identical percentiles for each of eight different distributions each with a mean of zero and a second moment of unity.

One thousand random observations from each distribution are provided, and these are arranged in twenty groups of fifty. The basic data for these have been drawn from the first two pages of Wold's (1954) *Random Normal Deviates*, and are given in the first column, x_1 , of each page. The further columns have been derived as follows.

(a). x_2 , random rectangular deviates from the distribution

$$f(x_2)dx_2 = \frac{1}{2\sqrt{3}}dx_2, \quad -\sqrt{3} \leq x_2 \leq \sqrt{3}. \quad (1)$$

These were obtained by transforming the normal deviates, x_1 , by

$$x_2 = 2\sqrt{3} \left[\int_{-\infty}^{x_1} \frac{1}{\sqrt{(2\pi)}} \exp -\frac{1}{2}x_1^2 dx_1 - \frac{1}{2} \right].$$

In this calculation, the expression in the bracket was carried to four decimal places and multiplied by 3.4641. The answer was rounded off to two decimal places.

(b). x_3 , random deviates from a distribution whose logarithm was normally distributed

$$f(x_3) dx_3 = \frac{1}{\sqrt{(2\pi)(x_3+a)}} \exp -\frac{1}{2} [\log_e (x_3+a) + b]^2 dx_3, \quad (x_3 \geq -a) \quad (2)$$

where $a = (e-1)^{-\frac{1}{2}} = 0.76287$ and $b = 0.5 - \log_e a = 0.771$. These were obtained by transforming the normal deviates, x_1 , by

$$x_3 = (e^{x_1} - e^{\frac{1}{2}}) / \sqrt{(e^2 - e)} = 0.46271 e^{x_1} - 0.76287.$$

This gives a lower limit for x_3 of 0.763 and a modal value of -0.593 . In making this transformation, e^{x_1} was calculated to three decimal places and the value of x_3 rounded off to two decimal places.

(c). x_4 , random deviates from the exponential distribution

$$f(x_4) dx_4 = e^{-(x_4+1)} dx_4 \quad (x_4 \geq -1). \quad (3)$$

These were obtained by transforming the normal deviates, x_1 , by

$$x_4 = -\log_e \left[\int_{x_1}^{\infty} \frac{1}{\sqrt{(2\pi)}} \exp -\frac{1}{2} x_1^2 dx_1 \right] - 1.$$

(d). x_5, x_6, x_7 , random deviates from an Edgeworth Type A expansion

$$f(x) dx = \exp \left[-\frac{\kappa_3}{3!} D^3 + \frac{\kappa_4}{4!} D^4 \right] \frac{1}{\sqrt{(2\pi)}} \exp -\frac{1}{2} x^2 dx, \quad (4)$$

with x_5 : $\kappa_3 = 0.5$, $\kappa_4 = 0.0$, with x_6 : $\kappa_3 = 0.0$, $\kappa_4 = 0.5$ and with x_7 : $\kappa_3 = 0.5$, $\kappa_4 = 0.5$.

These have been calculated using Cornish-Fisher expansions which give a sufficient degree of approximation within the central range of observations*. The expansions used, derived by substitution in formula (6.54) of Kendall and Stuart (1958), were

$$124,416x_5 = -9552 + 127,225x_1 + 7824x_1^2 - 40x_1^3 + 576x_1^4 - 252x_1^5,$$

$$1536x_6 = 1411x_1 + 56x_1^3 - 3x_1^5,$$

$$124,416x_7 = -12,144 + 122,878x_1 + 14,304x_1^2 - 1066x_1^3 - 720x_1^4 + 261x_1^5.$$

Tabulated values of these functions are given on page 182.

It is clear that these (or any other polynomial) expansions will break down for sufficiently extreme values of x_1 and the fact that tabulated values of x_5 reach a minimum near $x_1 = -2.5$ indicates the unreliability of the series expansion for x_5 beyond this point. To preserve the monotonic relationship between x_1 and x_5 , the values of x_5 corresponding to three more extreme values of x_1 (-2.79 , -2.76 , -2.69) have been replaced by the minimum value for x_5 , -1.86 . These replacements have been indicated with an asterisk. The original values, should they be needed, are respectively -1.81 , -1.82 , -1.84 .

A further limitation of the polynomial expansions is that values of cumulants of the distribution will not exactly equal the values of κ_3 and κ_4 inserted in the Edgeworth expansion. Thus the true values of the third and fourth cumulants (evaluated by taking expectations of powers of the polynomial expansions) are:

	x_5	x_6	x_7
Third cumulant	0.47	0.00	0.50
Fourth cumulant	0.01	0.45	0.56

* See note by E. S. Pearson on pp. 203-4 below.

(e). x_8 , random observations from the two-sided exponential distribution

$$f(x_8) dx_8 = \frac{1}{2\sqrt{2}} \exp\left(-\frac{|x_8|}{\sqrt{2}}\right) dx_8 \quad (-\infty \leq x_8 \leq \infty). \quad (5)$$

These were obtained by transforming the normal deviates, x_1 , by

$$x_8 = \frac{-\text{sign } x_1}{\sqrt{2}} \left[\log_e \int_{|x_1|}^{\infty} \frac{1}{\sqrt{(2\pi)}} \exp -\frac{1}{2}x_1^2 dx_1 + \log_e 2 \right].$$

The values were calculated to four decimal places and rounded off to two decimal places.

All calculations were repeated (in many instances by the same computer) to provide a check. A further check was provided by punching each row of results on a Hollerith card and ordering the set of these according to values of x_1 . All other variables should then be arranged in order.

This check revealed (apart from a number of errors in punching the cards) four or five omissions of minus signs. No gross error was found.

PROPERTIES OF THE DISTRIBUTIONS

The theoretical means and variances were, of course, zero and unity. The overall observed values showed close agreement with these theoretical values:

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Mean	0.00	0.00	-0.02	-0.01	0.00	0.00	0.00	0.00
Second moment about origin	0.95	0.98	0.81*	0.93	0.96	0.95	0.95	0.93

Theoretical values of the cumulants of the different distributions are shown in the following table:

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
κ_3	0.0	0.0	6.2	2.0	0.5	0.0	0.5	0.0
κ_4	0.0	-1.2	110.9	6.0	0.0	0.5	0.6	3.0
κ_5	0.0	0.0	5154.7	24.0	0.0	0.0	0.0	0.0
κ_6	0.0	6.9	617375	120.0	0.0	0.0	0.0	30.0

The variations in the forms of distributions are fairly well demonstrated by the following frequency table for the whole group of observations:

* Owing to the high value of κ_4 for this distribution, this seemingly extreme value corresponds to a two-tailed probability of greater than 0.5.

Values of the transformations for the Edgeworth Type A distributions

	x_1	x_2^*	x_3	x_4	x_5	x_6	x_7	x_8
8.00 to 8.99	.	.	1	.	.	.	.	.
7.00 to 7.99	.	.	2	.	.	.	.	.
6.00 to 6.99	.	.	1	.	.	.	.	.
5.00 to 5.99	.	.	0	3	.	.	.	1
4.00 to 4.99	.	.	5	2	.	.	1	3
3.60 to 3.99	.	.	2	4	.	1	3	0
3.20 to 3.59	.	.	5	7	3	3	1	5
2.80 to 3.19	3	.	3	4	2	7	13	8
2.40 to 2.79	3	.	4	5	11	11	8	7
2.00 to 2.39	14	.	5	23	10	27	33	24
1.60 to 1.99	29	34	20	18	33	58	58	40
1.20 to 1.59	62	117	18	43	60	86	79	70
0.80 to 1.19	100	122	43	59	87	136	116	124
0.40 to 0.79	131	117	70	78	120	169	146	216
0.00 to 0.39	156	108	128	121	145	162	167	210
-0.40 to -0.01	152	120	296	182	154	154	174	152
-0.80 to -0.41	150	110	397	293	154	86	112	55
-1.20 to -0.81	96	141	.	158	117	44	56	32
-1.60 to -1.21	48	95	.	.	66	27	26	21
-2.00 to -1.61	30	36	.	.	38	17	5	12
-2.40 to -2.01	20	.	.	.	.	9	2	12
-2.80 to -2.41	6	.	.	.	.	3	.	5
-3.20 to -2.81	.	.	.	.	.	.	.	1
-3.60 to -3.21	.	.	.	.	.	.	.	.
-4.00 to -3.61	.	.	.	.	.	.	.	.

* The expected numbers in the middle eight groupings here are 115.47. The two end groups have expected numbers 38.12.

REFERENCES

- FIELLER, E. C., LEWIS, T. & PEARSON, E. S. (1955). *Random Correlated Normal Deviates*. Cambridge University Press.
- KENDALL, M. G. (1949). Tables of autoregressive series. *Biometrika*, **36**, 267-89.
- KENDALL, M. G. & BABINGTON SMITH, B. (1939). *Tables of Random Sampling Numbers*. Cambridge University Press.
- KENDALL, M. G. & STUART, A. (1958). *The Advanced Theory of Statistics*, vol. 1. Charles Griffin and Co. Rand Corporation (1955). *A Million Random Digits with 100,000 Normal Deviates*. The Free Press, Illinois.
- TIPPETT, L. H. C. (1927). *Random Sampling Numbers*. Cambridge University Press.
- WOLD, H. (1954). *Random Normal Deviates*. Cambridge University Press.

$$(a) x_5 = \frac{1}{124416} [-9552 + 127,225x + 7824x^2 - 40x^3 + 576x^4 - 252x^5]$$

x	x_5	x	x_5	x	x_5	x	x_5	x	x_5
-3.2	↑	-1.9	-1.680	-0.6	-0.667	0.7	0.670	2.0	2.227
-3.1		-1.8	-1.625	-0.5	-0.572	0.8	0.782	2.1	2.352
-3.0		-1.7	-1.564	-0.4	-0.476	0.9	0.896	2.2	2.478
-2.9		-1.6	-1.499	-0.3	-0.378	1.0	1.011	2.3	2.603
-2.8		-1.5	-1.429	-0.2	-0.279	1.1	1.127	2.4	2.728
-2.7		-1.4	-1.356	-0.1	-0.178	1.2	1.245	2.5	2.851
-2.6		-1.3	-1.278	0	-0.077	1.3	1.364	2.6	2.972
-2.5	-1.857	-1.2	-1.198	0.1	0.026	1.4	1.484	2.7	3.092
-2.4	-1.849	-1.1	-1.115	0.2	0.130	1.5	1.606	2.8	3.208
-2.3	-1.832	-1.0	-1.029	0.3	0.236	1.6	1.728	2.9	3.322
-2.2	-1.806	-0.9	-0.942	0.4	0.342	1.7	1.852	3.0	3.431
-2.1	-1.771	-0.8	-0.852	0.5	0.450	1.8	1.976	3.1	3.536
-2.0	-1.729	-0.7	-0.760	0.6	0.560	1.9	2.101	3.2	3.635

$$(b) x_6 = \frac{1}{1536} [1411x + 56x^3 - 3x^5]$$

$\pm x$	$\pm x_6$	$\pm x$	$\pm x_6$	$\pm x$	$\pm x_6$	$\pm x$	$\pm x_6$	$\pm x$	$\pm x_6$
0.0	0.000	0.7	0.655	1.4	1.376	2.1	2.187	2.7	2.918
0.1	0.092	0.8	0.753	1.5	1.486	2.2	2.309	2.8	3.036
0.2	0.184	0.9	0.852	1.6	1.599	2.3	2.431	2.9	3.153
0.3	0.277	1.0	0.953	1.7	1.713	2.4	2.553	3.0	3.266
0.4	0.370	1.1	1.056	1.8	1.829	2.5	2.675	3.1	3.375
0.5	0.464	1.2	1.160	1.9	1.947	2.6	2.797	3.2	3.479
0.6	0.559	1.3	1.267	2.0	2.066	.	.	.	.

N.B. Negative values of x give negative values of x_6 .

$$(c) x_7 = \frac{1}{124,416} [-12,144 + 122,878x + 14,304x^2 - 1066x^3 - 720x^4 + 261x^5]$$

x	x_7	x	x_7	x	x_7	x	x_7	x	x_7
-3.2	-3.111	-1.9	-1.628	-0.6	-0.648	0.7	0.646	2.0	2.244
-3.1	-2.934	-1.8	-1.553	-0.5	-0.562	0.8	0.760	2.1	2.377
-3.0	-2.773	-1.7	-1.480	-0.4	-0.474	0.9	0.876	2.2	2.513
-2.9	-2.625	-1.6	-1.408	-0.3	-0.383	1.0	0.993	2.3	2.651
-2.8	-2.490	-1.5	-1.337	-0.2	-0.290	1.1	1.111	2.4	2.792
-2.7	-2.366	-1.4	-1.265	-0.1	-0.195	1.2	1.232	2.5	2.935
-2.6	-2.251	-1.3	-1.193	0	-0.098	1.3	1.353	2.6	3.082
-2.5	-2.145	-1.2	-1.120	0.1	0.002	1.4	1.476	2.7	3.232
-2.4	-2.046	-1.1	-1.045	0.2	0.104	1.5	1.600	2.8	3.386
-2.3	-1.954	-1.0	-0.970	0.3	0.209	1.6	1.726	2.9	3.545
-2.2	-1.866	-0.9	-0.892	0.4	0.315	1.7	1.853	3.0	3.710
-2.1	-1.784	-0.8	-0.813	0.5	0.424	1.8	1.982	3.1	3.880
-2.0	-1.704	-0.7	-0.731	0.6	0.534	1.9	2.112	3.2	4.056

First fifty

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	-0.84	-1.04	-0.56	-0.78	-0.89	-0.79	-0.84	-0.65
2	1.37	1.44	1.06	1.46	1.45	1.34	1.44	1.25
3	-0.18	-0.25	-0.38	-0.44	-0.26	-0.17	-0.27	-0.11
4	0.35	0.47	-0.11	0.01	0.29	0.32	0.26	0.23
5	2.82	1.72	7.00	5.03	3.23	3.06	3.42	3.78
6	2.12	1.67	3.09	3.07	2.38	2.21	2.40	2.39
7	-0.99	-1.17	-0.59	-0.82	-1.02	-0.94	-0.96	-0.80
8	0.34	0.46	-0.11	0.00	0.28	0.31	0.25	0.22
9	0.64	0.83	0.11	0.34	0.60	0.60	0.58	0.46
10	0.11	0.15	-0.25	-0.22	0.04	0.10	0.01	0.06
11	-0.02	-0.03	-0.31	-0.32	-0.10	-0.02	-0.12	-0.01
12	-0.86	-1.06	-0.57	-0.78	-0.91	-0.81	-0.86	-0.67
13	0.01	0.01	-0.30	-0.30	-0.07	0.01	-0.09	0.01
14	-0.03	-0.04	-0.31	-0.33	-0.11	-0.03	-0.13	-0.02
15	-0.41	-0.55	-0.46	-0.58	-0.49	-0.38	-0.48	-0.27
16	0.62	0.81	0.10	0.32	0.58	0.58	0.56	0.44
17	0.83	1.03	0.30	0.59	0.82	0.78	0.79	0.64
18	-0.78	-0.98	-0.55	-0.75	-0.83	-0.73	-0.80	-0.59
19	0.03	0.04	-0.29	-0.28	-0.05	0.03	-0.07	0.02
20	-2.30	-1.70	-0.72	-0.99	-1.83	-2.43	-1.95	-2.72
21	0.78	0.98	0.25	0.52	0.76	0.73	0.74	0.59
22	0.66	0.85	0.13	0.37	0.63	0.62	0.60	0.48
23	-0.34	-0.46	-0.43	-0.54	-0.42	-0.31	-0.42	-0.22
24	1.16	1.31	0.71	1.10	1.20	1.12	1.18	0.99
25	0.76	0.96	0.23	0.50	0.74	0.71	0.71	0.57
26	-1.00	-1.18	-0.59	-0.83	-1.03	-0.95	-0.97	-0.81
27	-0.32	-0.43	-0.43	-0.53	-0.40	-0.30	-0.40	-0.20
28	0.18	0.25	-0.21	-0.15	0.11	0.17	0.08	0.11
29	0.08	0.11	-0.26	-0.24	0.01	0.07	-0.02	0.05
30	1.98	1.65	2.59	2.73	2.20	2.04	2.22	2.15
31	0.91	1.10	0.39	0.71	0.91	0.86	0.89	0.72
32	-0.32	-0.43	-0.43	-0.53	-0.40	-0.30	-0.40	-0.20
33	-1.77	-1.60	-0.68	-0.96	-1.61	-1.79	-1.53	-1.81
34	0.16	0.22	-0.22	-0.17	0.09	0.15	0.06	0.10
35	0.79	0.99	0.26	0.54	0.77	0.74	0.75	0.60
36	0.20	0.27	-0.20	-0.13	0.13	0.18	0.10	0.12
37	1.31	1.40	0.95	1.35	1.38	1.28	1.37	1.17
38	0.35	0.47	-0.11	0.01	0.29	0.32	0.26	0.23
39	0.74	0.94	0.21	0.47	0.72	0.69	0.69	0.55
40	0.27	0.37	-0.16	-0.07	0.20	0.25	0.18	0.17
41	1.39	1.45	1.09	1.50	1.47	1.37	1.46	1.28
42	-0.14	-0.19	-0.36	-0.41	-0.22	-0.13	-0.23	-0.08
43	-0.76	-0.96	-0.55	-0.75	-0.82	-0.71	-0.78	-0.57
44	0.37	0.50	-0.09	0.03	0.31	0.34	0.28	0.24
45	-2.28	-1.69	-0.72	-0.99	-1.83	-2.41	-1.94	-2.68
46	0.36	0.49	-0.10	0.02	0.30	0.33	0.27	0.23
47	1.37	1.44	1.06	1.46	1.45	1.34	1.44	1.25
48	1.04	1.22	0.55	0.90	1.06	0.99	1.04	0.86
49	-0.67	-0.86	-0.53	-0.71	-0.73	-0.63	-0.71	-0.49
50	0.39	0.53	-0.08	0.05	0.33	0.36	0.30	0.25
$\Sigma(x)$	10.48	11.51	8.42	9.48	10.71	10.17	10.36	9.31
$\Sigma(x^2)$	51.4936	46.6287	76.3842	63.9778	53.0095	52.9419	53.8870	56.9311

Second fifty

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	0.68	0.87	0.15	0.39	0.65	0.64	0.62	0.49
2	-0.77	-0.97	-0.55	-0.75	-0.82	-0.72	-0.79	-0.58
3	-1.39	-1.45	-0.65	-0.91	-1.35	-1.37	-1.26	-1.28
4	-1.91	-1.63	-0.69	-0.97	-1.68	-1.96	-1.64	-2.04
5	0.05	0.07	-0.28	-0.27	-0.03	0.05	-0.05	0.03
6	1.32	1.41	0.97	1.37	1.39	1.29	1.38	1.19
7	-0.67	-0.86	-0.53	-0.71	-0.73	-0.63	-0.71	-0.49
8	-0.81	-1.01	-0.56	-0.77	-0.86	-0.76	-0.82	-0.62
9	0.96	1.15	0.45	0.78	0.97	0.91	0.95	0.77
10	-0.24	-0.33	-0.40	-0.48	-0.32	-0.22	-0.33	-0.15
11	-0.66	-0.85	-0.52	-0.71	-0.72	-0.62	-0.70	-0.48
12	1.08	1.25	0.60	0.97	1.10	1.04	1.09	0.90
13	0.27	0.37	-0.16	-0.07	0.20	0.25	0.18	0.17
14	0.46	0.61	-0.03	0.13	0.41	0.43	0.38	0.31
15	-0.77	-0.97	-0.55	-0.75	-0.82	-0.72	-0.79	-0.58
16	0.00	0.00	-0.30	-0.31	-0.08	0.00	-0.10	0.00
17	1.01	1.19	0.51	0.86	1.02	0.96	1.00	0.82
18	-0.92	-1.11	-0.58	-0.80	-0.96	-0.87	-0.91	-0.73
19	0.98	1.17	0.47	0.81	0.99	0.93	0.97	0.79
20	0.24	0.33	-0.17	-0.10	0.17	0.22	0.15	0.15
21	1.01	1.19	0.51	0.86	1.02	0.96	1.00	0.82
22	1.49	1.50	1.29	1.69	1.59	1.48	1.59	1.41
23	1.16	1.31	0.71	1.10	1.20	1.12	1.18	0.99
24	0.48	0.64	-0.02	0.15	0.43	0.45	0.40	0.33
25	0.40	0.54	-0.07	0.07	0.34	0.37	0.32	0.26
26	-0.12	-0.17	-0.35	-0.40	-0.20	-0.11	-0.21	-0.07
27	-0.85	-1.05	-0.57	-0.78	-0.90	-0.80	-0.85	-0.66
28	-0.46	-0.61	-0.47	-0.61	-0.53	-0.43	-0.53	-0.31
29	0.35	0.47	-0.11	0.01	0.29	0.32	0.26	0.23
30	0.23	0.32	-0.18	-0.11	0.16	0.21	0.14	0.14
31	0.20	0.27	-0.20	-0.13	0.13	0.18	0.10	0.12
32	-0.17	-0.23	-0.37	-0.43	-0.25	-0.16	-0.26	-0.10
33	0.12	0.17	-0.24	-0.21	0.05	0.11	0.02	0.07
34	-1.48	-1.49	-0.66	-0.93	-1.41	-1.46	-1.32	-1.40
35	0.52	0.69	0.02	0.20	0.47	0.48	0.45	0.36
36	1.56	1.53	1.44	1.82	1.68	1.55	1.68	1.51
37	1.04	1.22	0.55	0.90	1.06	0.99	1.04	0.86
38	-0.18	-0.25	-0.38	-0.44	-0.26	-0.17	-0.27	-0.11
39	-0.26	-0.36	-0.41	-0.49	-0.34	-0.24	-0.35	-0.16
40	1.00	1.18	0.49	0.84	1.01	0.95	0.99	0.81
41	-0.51	-0.68	-0.49	-0.64	-0.58	-0.47	-0.57	-0.35
42	0.05	0.07	-0.28	-0.27	-0.03	0.05	-0.05	0.03
43	-0.69	-0.88	-0.53	-0.72	-0.75	-0.65	-0.72	-0.50
44	-0.58	-0.76	-0.50	-0.67	-0.65	-0.54	-0.63	-0.41
45	-0.69	-0.88	-0.53	-0.72	-0.75	-0.65	-0.72	-0.50
46	-1.41	-1.46	-0.65	-0.92	-1.36	-1.39	-1.27	-1.30
47	0.24	0.33	-0.17	-0.10	0.17	0.22	0.15	0.15
48	0.41	0.55	-0.07	0.08	0.35	0.38	0.33	0.27
49	0.15	0.21	-0.23	-0.18	0.08	0.14	0.05	0.09
50	-0.24	-0.33	-0.40	-0.48	-0.32	-0.22	-0.33	-0.15
$\Sigma(x)$	1.68	2.28	-5.69	-3.80	0.23	1.52	0.24	1.10
$\Sigma(x^2)$	32.9330	40.9668	13.6407	26.0324	33.5593	31.0336	31.6978	25.2532

Third fifty

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	0.11	0.15	-0.25	-0.22	0.04	0.10	0.01	0.06
2	-0.89	-1.09	-0.57	-0.79	-0.93	-0.84	-0.88	-0.70
3	0.99	1.17	0.48	0.83	1.00	0.94	0.98	0.80
4	0.50	0.66	0.00	0.18	0.45	0.46	0.42	0.34
5	-0.95	-1.14	-0.58	-0.81	-0.99	-0.90	-0.93	-0.76
6	0.81	1.01	0.28	0.57	0.79	0.76	0.77	0.62
7	-0.28	-0.38	-0.41	-0.51	-0.36	-0.26	-0.36	-0.18
8	-1.64	-1.56	-0.67	-0.95	-1.53	-1.64	-1.44	-1.62
9	-1.44	-1.47	-0.65	-0.92	-1.39	-1.42	-1.29	-1.34
10	0.01	0.01	-0.30	-0.30	-0.07	0.01	-0.09	0.01
11	0.24	0.33	-0.17	-0.10	0.17	0.22	0.15	0.15
12	0.46	0.61	-0.03	0.13	0.41	0.43	0.38	0.31
13	0.02	0.03	-0.29	-0.29	-0.06	0.02	-0.08	0.01
14	-2.09	-1.67	-0.71	-0.98	-1.77	-2.17	-1.78	-2.34
15	2.08	1.67	2.94	2.97	2.33	2.16	2.35	2.32
16	-0.80	-1.00	-0.56	-0.76	-0.85	-0.75	-0.81	-0.61
17	0.27	0.37	-0.16	-0.07	0.20	0.25	0.18	0.17
18	0.21	0.29	-0.19	-0.12	0.14	0.19	0.11	0.13
19	-1.33	-1.41	-0.64	-0.90	-1.30	-1.30	-1.21	-1.20
20	1.70	1.58	1.77	2.11	1.85	1.71	1.85	1.71
21	-0.92	-1.11	-0.58	-0.80	-0.96	-0.87	-0.91	-0.73
22	1.73	1.59	1.85	2.17	1.89	1.75	1.89	1.75
23	0.66	0.85	0.13	0.37	0.63	0.62	0.60	0.48
24	-1.69	-1.57	-0.68	-0.95	-1.56	-1.70	-1.47	-1.19
25	2.13	1.67	3.13	3.10	2.39	2.22	2.42	2.41
26	-2.20	-1.68	-0.71	-0.99	-1.81	-2.31	-1.87	-2.53
27	1.57	1.53	1.46	1.84	1.69	1.57	1.69	1.52
28	-1.85	-1.62	-0.69	-0.97	-1.65	-1.89	-1.59	-1.94
29	-2.24	-1.69	-0.71	-0.99	-1.82	-2.36	-1.90	-2.61
30	-0.15	-0.21	-0.36	-0.42	-0.23	-0.14	-0.24	-0.09
31	-0.45	-0.60	-0.47	-0.60	-0.52	-0.42	-0.52	-0.30
32	-0.03	-0.04	-0.31	-0.33	-0.11	-0.03	-0.13	-0.02
33	1.01	1.19	0.51	0.86	1.02	0.96	1.00	0.82
34	1.19	1.33	0.76	1.15	1.23	1.15	1.22	1.03
35	-1.03	-1.21	-0.60	-0.84	-1.05	-0.98	-0.99	-0.84
36	0.35	0.47	-0.11	0.01	0.29	0.32	0.26	0.23
37	-0.82	-1.02	-0.56	-0.77	-0.87	-0.77	-0.83	-0.63
38	1.76	1.60	1.93	2.24	1.93	1.78	1.93	1.80
39	-0.68	-0.87	-0.53	-0.71	-0.74	-0.64	-0.71	-0.49
40	-0.88	-1.08	-0.57	-0.79	-0.92	-0.83	-0.88	-0.69
41	-0.14	-0.19	-0.36	-0.41	-0.22	-0.13	-0.23	-0.08
42	0.77	0.97	0.24	0.51	0.75	0.72	0.73	0.58
43	-1.61	-1.55	-0.67	-0.94	-1.51	-1.61	-1.42	-1.58
44	-0.67	-0.86	-0.53	-0.71	-0.73	-0.63	-0.71	-0.49
45	0.58	0.76	0.06	0.27	0.54	0.54	0.51	0.41
46	-1.06	-1.23	-0.60	-0.84	-1.08	-1.01	-1.02	-0.88
47	0.51	0.68	0.01	0.19	0.46	0.47	0.44	0.35
48	-0.24	-0.33	-0.40	-0.48	-0.32	-0.22	-0.33	-0.15
49	-1.60	-1.54	-0.67	-0.94	-1.50	-1.60	-1.41	-1.56
50	0.29	0.40	-0.14	-0.05	0.23	0.27	0.20	0.18
$\Sigma(x)$	-7.73	-7.20	-0.88	-1.75	-6.42	-7.80	-5.94	-7.36
$\Sigma(x^2)$	67.1551	62.3312	41.2024	56.4089	65.3910	67.9222	63.2752	66.3356

Fourth fifty

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	1.05	1.22	0.56	0.92	1.07	1.00	1.05	0.87
2	0.61	0.79	0.09	0.31	0.57	0.57	0.55	0.43
3	-0.01	-0.01	-0.30	-0.31	-0.09	-0.01	-0.11	-0.01
4	-0.92	-1.11	-0.58	-0.80	-0.96	-0.87	-0.91	-0.73
5	1.17	1.31	0.73	1.11	1.21	1.13	1.20	1.00
6	-1.91	-1.63	-0.69	-0.97	-1.68	-1.96	-1.64	-2.04
7	-0.48	-0.64	-0.48	-0.62	-0.55	-0.45	-0.54	-0.33
8	-0.21	-0.29	-0.39	-0.46	-0.29	-0.19	-0.30	-0.13
9	-2.47	-1.71	-0.72	-0.99	-1.85	-2.64	-2.12	-3.04
10	-0.28	-0.38	-0.41	-0.51	-0.36	-0.26	-0.36	-0.18
11	-0.43	-0.58	-0.46	-0.59	-0.50	-0.40	-0.50	-0.29
12	0.08	0.11	-0.26	-0.24	0.01	0.07	-0.02	0.05
13	-0.60	-0.78	-0.51	-0.68	-0.67	-0.56	-0.65	-0.42
14	-1.60	-1.54	-0.67	-0.94	-1.50	-1.60	-1.41	-1.56
15	-0.98	-1.17	-0.59	-0.82	-1.01	-0.93	-0.95	-0.79
16	-0.62	-0.81	-0.51	-0.69	-0.69	-0.58	-0.66	-0.44
17	1.47	1.49	1.25	1.65	1.57	1.45	1.56	1.38
18	-0.60	-0.78	-0.51	-0.68	-0.67	-0.56	-0.65	-0.42
19	0.66	0.85	0.13	0.37	0.63	0.62	0.60	0.48
20	-0.52	-0.69	-0.49	-0.64	-0.59	-0.48	-0.58	-0.36
21	1.82	1.61	2.09	2.37	2.00	1.85	2.01	1.89
22	-0.27	-0.37	-0.41	-0.50	-0.35	-0.25	-0.36	-0.17
23	0.59	0.77	0.07	0.28	0.55	0.55	0.52	0.42
24	0.07	0.10	-0.27	-0.25	0.00	0.06	-0.03	0.04
25	-0.25	-0.34	-0.40	-0.49	-0.33	-0.23	-0.34	-0.16
26	0.75	0.95	0.22	0.48	0.73	0.70	0.70	0.56
27	0.25	0.34	-0.17	-0.09	0.18	0.23	0.16	0.16
28	2.24	1.69	3.58	3.38	2.53	2.36	2.57	2.61
29	-0.27	-0.37	-0.41	-0.50	-0.35	-0.25	-0.36	-0.17
30	-0.22	-0.30	-0.39	-0.47	-0.30	-0.20	-0.31	-0.14
31	0.06	0.08	-0.27	-0.26	-0.02	0.06	-0.04	0.03
32	-0.75	-0.95	-0.54	-0.74	-0.81	-0.70	-0.77	-0.56
33	-2.27	-1.69	-0.72	-0.99	-1.82	-2.39	-1.93	-2.66
34	-0.23	-0.32	-0.40	-0.47	-0.31	-0.21	-0.32	-0.14
35	-1.18	-1.32	-0.62	-0.87	-1.18	-1.14	-1.11	-1.02
36	-0.62	-0.81	-0.51	-0.69	-0.69	-0.58	-0.66	-0.44
37	0.47	0.63	-0.02	0.14	0.42	0.44	0.39	0.32
38	0.92	1.11	0.40	0.72	0.92	0.87	0.90	0.73
39	-0.32	-0.43	-0.43	-0.53	-0.40	-0.30	-0.40	-0.20
40	-0.23	-0.32	-0.40	-0.47	-0.31	-0.21	-0.32	-0.14
41	-0.89	-1.09	-0.57	-0.79	-0.93	-0.84	-0.88	-0.70
42	1.33	1.41	0.99	1.39	1.40	1.30	1.39	1.20
43	0.33	0.45	-0.12	-0.01	0.27	0.30	0.24	0.21
44	0.68	0.87	0.15	0.39	0.65	0.64	0.62	0.49
45	-0.38	-0.51	-0.45	-0.57	-0.46	-0.35	-0.46	-0.25
46	-0.89	-1.09	-0.57	-0.79	-0.93	-0.84	-0.88	-0.70
47	-0.28	-0.38	-0.41	-0.51	-0.36	-0.26	-0.36	-0.28
48	0.19	0.26	-0.20	-0.14	0.12	0.17	0.09	0.12
49	0.84	1.04	0.31	0.61	0.83	0.79	0.81	0.65
50	-0.03	-0.04	-0.31	-0.33	-0.11	-0.03	-0.13	-0.02
$\Sigma(x)$	-5.13	-5.37	-5.59	-6.28	-5.41	-5.11	-5.70	-4.85
$\Sigma(x^2)$	44.7007	43.2703	28.9523	39.2082	42.8955	45.2211	43.0940	45.6167

Fifth fifty

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	0.09	0.12	-0.26	-0.23	0.02	0.08	-0.01	0.05
2	-2.54	-1.71	-0.73	-0.99	-1.86	-2.72	-2.19	-3.19
3	-0.22	-0.30	-0.39	-0.47	-0.30	-0.20	-0.31	-0.14
4	-0.30	-0.41	-0.42	-0.52	-0.38	-0.28	-0.38	-0.19
5	1.22	1.35	0.80	1.20	1.27	1.18	1.26	1.06
6	-1.23	-1.35	-0.63	-0.88	-1.22	-1.19	-1.14	-1.07
7	-0.12	-0.17	-0.35	-0.40	-0.20	-0.11	-0.21	-0.07
8	-0.19	-0.26	-0.38	-0.45	-0.27	-0.17	-0.28	-0.12
9	-1.69	-1.57	-0.68	-0.95	-1.56	-1.70	-1.47	-1.69
10	0.04	0.06	-0.28	-0.27	-0.04	0.04	-0.06	0.02
11	1.09	1.25	0.61	0.98	1.12	1.05	1.10	0.91
12	1.03	1.21	0.53	0.89	1.05	0.98	1.03	0.84
13	-0.61	-0.79	-0.51	-0.68	-0.68	-0.57	-0.66	-0.43
14	-0.56	-0.74	-0.50	-0.66	-0.63	-0.52	-0.61	-0.39
15	0.36	0.49	-0.10	0.02	0.30	0.33	0.27	0.23
16	0.35	0.47	-0.11	0.01	0.29	0.32	0.26	0.23
17	0.29	0.40	-0.14	-0.05	0.23	0.27	0.20	0.18
18	-0.59	-0.77	-0.51	-0.67	-0.66	-0.55	-0.64	-0.42
19	-0.57	-0.75	-0.50	-0.67	-0.64	-0.53	-0.62	-0.40
20	-0.15	-0.21	-0.36	-0.42	-0.23	-0.14	-0.24	-0.09
21	-0.67	-0.86	-0.53	-0.71	-0.73	-0.63	-0.71	-0.49
22	-0.87	-1.07	-0.57	-0.79	-0.92	-0.82	-0.87	-0.68
23	0.35	0.47	-0.11	0.01	0.29	0.32	0.26	0.23
24	0.70	0.89	0.17	0.42	0.67	0.66	0.65	0.51
25	-0.15	-0.21	-0.36	-0.42	-0.23	-0.14	-0.24	-0.09
26	1.30	1.40	0.93	1.34	1.36	1.27	1.35	1.16
27	0.75	0.95	0.22	0.48	0.73	0.70	0.70	0.56
28	1.08	1.25	0.60	0.97	1.10	1.04	1.09	0.90
29	0.67	0.86	0.14	0.38	0.64	0.63	0.61	0.49
30	-0.56	-0.74	-0.50	-0.66	-0.63	-0.52	-0.61	-0.39
31	-1.10	-1.26	-0.61	-0.85	-1.12	-1.06	-1.05	-0.92
32	0.93	1.12	0.41	0.74	0.93	0.88	0.91	0.74
33	0.55	0.72	0.04	0.23	0.51	0.51	0.48	0.38
34	0.59	0.77	0.07	0.28	0.55	0.55	0.52	0.42
35	-0.85	-1.05	-0.57	-0.78	-0.90	-0.80	-0.85	-0.66
36	-0.15	-0.21	-0.36	-0.42	-0.23	-0.14	-0.24	-0.09
37	0.81	1.01	0.28	0.57	0.79	0.76	0.77	0.62
38	-1.21	-1.34	-0.62	-0.88	-1.21	-1.17	-1.13	-1.05
39	-0.44	-0.59	-0.46	-0.60	-0.51	-0.41	-0.51	-0.30
40	-0.47	-0.63	-0.47	-0.62	-0.54	-0.44	-0.54	-0.32
41	0.76	0.96	0.23	0.50	0.74	0.71	0.71	0.57
42	-0.58	-0.76	-0.50	-0.67	-0.65	-0.54	-0.63	-0.41
43	-0.13	-0.18	-0.36	-0.41	-0.21	-0.12	-0.22	-0.08
44	0.17	0.23	-0.21	-0.16	0.10	0.16	0.07	0.10
45	-1.65	-1.56	-0.67	-0.95	-1.53	-1.66	-1.44	-1.64
46	-0.28	-0.38	-0.41	-0.51	-0.36	-0.26	-0.36	-0.17
47	0.32	0.43	-0.13	-0.02	0.26	0.30	0.23	0.20
48	-2.10	-1.67	-0.71	-0.98	-1.77	-2.19	-1.78	-2.35
49	-0.07	-0.10	-0.33	-0.36	-0.15	-0.06	-0.17	-0.04
50	0.83	1.03	0.30	0.59	0.82	0.78	0.79	0.64
$\Sigma(x)$ $\Sigma(x^2)$	-5.77 37.4683	-4.20 40.9470	-10.00 10.6474	-9.49 21.3559	-6.59 33.6241	-6.12 36.8738	-6.91 33.0399	-6.84 34.6564

Sixth fifty

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	0.83	1.03	0.30	0.59	0.82	0.78	0.79	0.64
2	1.55	1.52	1.42	1.80	1.67	1.54	1.66	1.49
3	0.36	0.49	-0.10	0.02	0.30	0.33	0.27	0.23
4	0.36	0.49	-0.10	0.02	0.30	0.33	0.27	0.23
5	0.14	0.19	-0.23	-0.19	0.07	0.13	0.04	0.08
6	-0.17	-0.23	-0.37	-0.43	-0.25	-0.16	-0.26	-0.10
7	-2.21	-1.68	-0.71	-0.99	-1.81	-2.32	-1.87	-2.55
8	0.66	0.85	0.13	0.37	0.63	0.62	0.60	0.48
9	-0.45	-0.60	-0.47	-0.60	-0.52	-0.42	-0.52	-0.30
10	1.28	1.38	0.90	1.30	1.34	1.25	1.33	1.14
11	-0.41	-0.55	-0.46	-0.58	-0.49	-0.38	-0.48	-0.27
12	-1.73	-1.59	-0.68	-0.96	-1.58	-1.75	-1.50	-1.75
13	-1.01	-1.19	-0.59	-0.83	-1.04	-0.96	-0.98	-0.82
14	-1.38	-1.44	-0.65	-0.91	-1.34	-1.35	-1.25	-1.26
15	-0.36	-0.49	-0.44	-0.55	-0.44	-0.33	-0.44	-0.23
16	0.31	0.42	-0.13	-0.03	0.25	0.29	0.22	0.19
17	-2.79	-1.72	-0.73	-1.00	-1.86*	-3.02	-2.48	-3.72
18	0.14	0.19	-0.23	-0.19	0.07	0.13	0.04	0.08
19	-0.68	-0.87	-0.53	-0.71	-0.74	-0.64	-0.71	-0.49
20	1.13	1.28	0.67	1.05	1.16	1.09	1.15	0.96
21	-1.30	-1.40	-0.64	-0.90	-1.28	-1.27	-1.19	-1.16
22	-0.64	-0.83	-0.52	-0.70	-0.70	-0.60	-0.68	-0.46
23	-1.63	-1.55	-0.67	-0.95	-1.52	-1.63	-1.43	-1.61
24	-0.49	-0.65	-0.48	-0.63	-0.56	-0.45	-0.55	-0.33
25	-0.25	-0.34	-0.40	-0.49	-0.33	-0.23	-0.34	-0.16
26	-0.64	-0.83	-0.52	-0.70	-0.70	-0.60	-0.68	-0.46
27	-0.61	-0.79	-0.51	-0.68	-0.68	-0.57	-0.66	-0.43
28	0.24	0.33	-0.17	-0.10	0.17	0.22	0.15	0.15
29	-1.01	-1.19	-0.59	-0.83	-1.04	-0.96	-0.98	-0.82
30	-0.62	-0.81	-0.51	-0.69	-0.69	-0.58	-0.66	-0.44
31	-0.23	-0.32	-0.40	-0.47	-0.31	-0.21	-0.32	-0.14
32	0.42	0.56	-0.06	0.09	0.36	0.39	0.34	0.28
33	0.32	0.43	-0.13	-0.02	0.26	0.30	0.23	0.20
34	0.41	0.55	-0.07	0.08	0.35	0.38	0.33	0.27
35	1.67	1.57	1.70	2.05	1.81	1.68	1.81	1.66
36	-0.19	-0.26	-0.38	-0.45	-0.27	-0.17	-0.28	-0.12
37	0.26	0.36	-0.16	-0.08	0.19	0.24	0.17	0.16
38	-0.48	-0.64	-0.48	-0.62	-0.55	-0.45	-0.54	-0.33
39	0.01	0.01	-0.30	-0.30	-0.07	0.01	-0.09	0.01
40	-1.61	-1.55	-0.67	-0.94	-1.51	-1.61	-1.42	-1.58
41	0.14	0.19	-0.23	-0.19	0.07	0.13	0.04	0.08
42	-0.76	-0.96	-0.55	-0.75	-0.82	-0.71	-0.78	-0.57
43	0.30	0.41	-0.14	-0.04	0.24	0.28	0.21	0.19
44	-0.54	-0.71	-0.49	-0.65	-0.61	-0.50	-0.60	-0.37
45	-0.86	-1.06	-0.57	-0.78	-0.91	-0.81	-0.86	-0.67
46	-0.75	-0.95	-0.54	-0.74	-0.81	-0.70	-0.77	-0.56
47	0.32	0.43	-0.13	-0.02	0.26	0.30	0.23	0.20
48	-0.69	-0.88	-0.53	-0.72	-0.75	-0.65	-0.72	-0.50
49	-0.06	-0.08	-0.33	-0.35	-0.14	-0.06	-0.16	-0.03
50	0.67	0.86	0.14	0.38	0.64	0.63	0.61	0.49
$\Sigma(x)$	-13.03	-12.62	-12.33	-14.01	-13.36	-13.04	-13.71	-13.02
$\Sigma(x^2)$	43.2437	42.6990	15.1983	26.6345	38.1144	43.5442	38.7677	44.3438

Seventh fifty

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	1.35	1.43	1.02	1.42	1.42	1.32	1.41	1.22
2	-0.92	-1.11	-0.58	-0.80	-0.96	-0.87	-0.91	-0.73
3	1.27	1.38	0.88	1.28	1.33	1.23	1.32	1.12
4	1.28	1.38	0.90	1.30	1.34	1.25	1.33	1.14
5	-2.13	-1.67	-0.70	-0.98	-1.78	-2.22	-1.81	-2.41
6	-0.38	-0.51	-0.45	-0.57	-0.46	-0.35	-0.46	-0.25
7	-0.10	-0.14	-0.34	-0.38	-0.18	-0.09	-0.20	-0.06
8	0.12	0.17	-0.24	-0.21	0.05	0.11	0.02	0.07
9	2.31	1.70	3.90	3.57	2.62	2.44	2.67	2.74
10	0.47	0.63	-0.02	0.14	0.42	0.44	0.39	0.32
11	0.42	0.56	-0.06	0.09	0.36	0.39	0.34	0.28
12	-1.07	-1.24	-0.60	-0.85	-1.09	-1.03	-1.02	-0.89
13	0.22	0.30	-0.19	-0.12	0.15	0.20	0.13	0.14
14	-1.33	-1.41	-0.64	-0.90	-1.30	-1.30	-1.21	-1.20
15	-2.24	-1.69	-0.71	-0.99	-1.82	-2.36	-1.90	-2.61
16	-0.20	-0.27	-0.38	-0.45	-0.28	-0.18	-0.29	-0.12
17	0.20	0.27	-0.20	-0.13	0.13	0.18	0.10	0.12
18	0.09	0.12	-0.26	-0.23	0.02	0.08	-0.01	0.05
19	0.02	0.03	-0.29	-0.29	-0.06	0.02	-0.08	0.01
20	-0.23	-0.32	-0.40	-0.47	-0.31	-0.21	-0.32	-0.14
21	-1.07	-1.24	-0.60	-0.85	-1.09	-1.03	-1.02	-0.89
22	1.43	1.47	1.17	1.57	1.52	1.41	1.51	1.33
23	0.22	0.30	-0.19	-0.12	0.15	0.20	0.13	0.14
24	-1.24	-1.36	-0.63	-0.89	-1.23	-1.20	-1.15	-1.09
25	0.91	1.10	0.39	0.71	0.91	0.86	0.89	0.72
26	-1.88	-1.63	-0.69	-0.97	-1.67	-1.92	-1.61	-1.99
27	0.08	0.11	-0.26	-0.24	0.01	0.07	-0.02	0.05
28	2.19	1.68	3.37	3.25	2.47	2.30	2.50	2.51
29	0.17	0.23	-0.21	-0.16	0.10	0.16	0.07	0.10
30	-0.61	-0.79	-0.51	-0.68	-0.68	-0.57	-0.66	-0.43
31	-0.12	-0.17	-0.35	-0.40	-0.20	-0.11	-0.21	-0.07
32	0.70	0.89	0.17	0.42	0.67	0.66	0.65	0.51
33	-1.72	-1.58	-0.68	-0.96	-1.58	-1.74	-1.49	-1.74
34	-1.46	-1.48	-0.66	-0.93	-1.40	-1.44	-1.31	-1.37
35	-0.36	-0.49	-0.44	-0.55	-0.44	-0.33	-0.44	-0.23
36	-1.39	-1.45	-0.65	-0.91	-1.35	-1.37	-1.26	-1.28
37	0.01	0.01	-0.30	-0.30	-0.07	0.01	-0.09	0.01
38	0.52	0.69	0.02	0.20	0.47	0.48	0.45	0.36
39	-0.47	-0.63	-0.47	-0.62	-0.54	-0.44	-0.54	-0.32
40	-0.01	-0.01	-0.30	-0.31	-0.09	-0.01	-0.11	-0.01
41	-0.49	-0.65	-0.48	-0.63	-0.56	-0.45	-0.55	-0.33
42	0.19	0.26	-0.20	-0.14	0.12	0.17	0.09	0.12
43	-0.05	-0.07	-0.32	-0.35	-0.13	-0.05	-0.15	-0.03
44	1.07	1.24	0.59	0.95	1.09	1.03	1.08	0.89
45	-1.39	-1.45	-0.65	-0.91	-1.35	-1.37	-1.26	-1.28
46	0.96	1.15	0.45	0.78	0.97	0.91	0.95	0.77
47	1.32	1.41	0.97	1.37	1.39	1.29	1.38	1.19
48	-0.98	-1.17	-0.59	-0.82	-1.01	-0.93	-0.95	-0.79
49	-0.25	-0.34	-0.40	-0.49	-0.33	-0.23	-0.34	-0.16
50	-1.20	-1.33	-0.62	-0.88	-1.20	-1.16	-1.12	-1.54
$\Sigma(x)$	-5.77	-5.69	-2.43	-3.43	-5.45	-5.75	-5.08	-6.05
$\Sigma(x^2)$	55.4525	52.9583	40.6117	50.1883	55.0049	55.9079	53.1864	56.6803

Eighth fifty

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	1.01	1.19	0.51	0.86	1.02	0.96	1.00	0.82
2	-0.20	-0.27	-0.38	-0.45	-0.28	-0.18	-0.29	-0.12
3	0.45	0.60	-0.04	0.12	0.40	0.42	0.37	0.30
4	-0.69	-0.88	-0.53	-0.72	-0.75	-0.65	-0.72	-0.50
5	-0.76	-0.96	-0.55	-0.75	-0.82	-0.71	-0.78	-0.57
6	-0.03	-0.04	-0.31	-0.33	-0.11	-0.03	-0.13	-0.02
7	0.40	0.54	-0.07	0.07	0.34	0.37	0.32	0.26
8	0.39	0.53	-0.08	0.05	0.33	0.36	0.30	0.25
9	1.39	1.45	1.09	1.50	1.47	1.37	1.46	1.28
10	0.77	0.97	0.24	0.51	0.75	0.72	0.73	0.58
11	-0.37	-0.50	-0.44	-0.56	-0.45	-0.34	-0.45	-0.24
12	0.35	0.47	-0.11	0.01	0.29	0.32	0.26	0.23
13	-1.25	-1.37	-0.63	-0.89	-1.24	-1.21	-1.16	-1.10
14	-0.70	-0.89	-0.53	-0.72	-0.76	-0.66	-0.73	-0.51
15	1.41	1.46	1.13	1.53	1.50	1.39	1.49	1.30
16	-0.04	-0.06	-0.32	-0.34	-0.12	-0.04	-0.14	-0.02
17	0.42	0.56	-0.06	0.09	0.36	0.39	0.34	0.28
18	-0.09	-0.12	-0.34	-0.38	-0.17	-0.08	-0.19	-0.05
19	-0.59	-0.77	-0.51	-0.67	-0.66	-0.55	-0.64	-0.42
20	0.77	0.97	0.24	0.51	0.75	0.72	0.73	0.58
21	-0.69	-0.88	-0.53	-0.72	-0.75	-0.65	-0.72	-0.50
22	0.14	0.19	-0.23	-0.19	0.07	0.13	0.04	0.08
23	-0.34	-0.46	-0.43	-0.54	-0.42	-0.31	-0.42	-0.22
24	-0.51	-0.68	-0.49	-0.64	-0.58	-0.47	-0.57	-0.35
25	0.54	0.71	0.03	0.22	0.49	0.50	0.47	0.37
26	1.35	1.43	1.02	1.42	1.42	1.32	1.41	1.22
27	0.06	0.08	-0.27	-0.26	-0.02	0.06	-0.04	0.03
28	0.66	0.85	0.13	0.37	0.63	0.62	0.60	0.48
29	-0.35	-0.47	-0.44	-0.55	-0.43	-0.32	-0.43	-0.23
30	-0.53	-0.70	-0.49	-0.65	-0.60	-0.49	-0.59	-0.37
31	-0.56	-0.74	-0.50	-0.66	-0.63	-0.52	-0.61	-0.39
32	-0.02	-0.03	-0.31	-0.32	-0.10	-0.02	-0.12	-0.01
33	0.66	0.85	0.13	0.37	0.63	0.62	0.60	0.48
34	0.10	0.14	-0.25	-0.22	0.03	0.09	0.00	0.06
35	0.52	0.69	0.02	0.20	0.47	0.48	0.45	0.36
36	-0.46	-0.61	-0.47	-0.61	-0.53	-0.43	-0.53	-0.31
37	0.47	0.63	-0.02	0.14	0.42	0.44	0.39	0.32
38	0.40	0.54	-0.07	0.07	0.34	0.37	0.32	0.26
39	-1.15	-1.30	-0.62	-0.87	-1.16	-1.11	-1.08	-0.98
40	-0.29	-0.40	-0.42	-0.51	-0.37	-0.27	-0.37	-0.18
41	0.82	1.02	0.29	0.58	0.81	0.77	0.78	0.63
42	0.23	0.32	-0.18	-0.11	0.16	0.21	0.14	0.14
43	-0.87	-1.07	-0.57	-0.79	-0.92	-0.82	-0.87	-0.68
44	-0.12	-0.17	-0.35	-0.40	-0.20	-0.11	-0.21	-0.07
45	0.64	0.83	0.11	0.34	0.60	0.60	0.58	0.46
46	-0.03	-0.04	-0.31	-0.33	-0.11	-0.03	-0.13	-0.02
47	0.17	0.23	-0.21	-0.16	0.10	0.16	0.07	0.10
48	1.49	1.50	1.29	1.69	1.59	1.48	1.59	1.41
49	-0.51	-0.68	-0.49	-0.64	-0.58	-0.47	-0.57	-0.35
50	0.62	0.81	0.10	0.32	0.58	0.58	0.56	0.44
$\Sigma(x)$	5.08	5.47	-6.22	-4.01	2.79	4.98	2.51	4.51
$\Sigma(x^2)$	22.2392	30.8785	11.2092	20.7837	23.8075	20.3996	22.5705	14.9801

Ninth fifty

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	-0.54	-0.71	-0.49	-0.65	-0.61	-0.50	-0.60	-0.37
2	1.11	1.27	0.64	1.01	1.14	1.07	1.12	0.94
3	0.59	0.77	0.07	0.28	0.55	0.55	0.52	0.42
4	-1.64	-1.56	-0.67	-0.95	-1.53	-1.64	-1.44	-1.62
5	0.24	0.33	-0.17	-0.10	0.17	0.22	0.15	0.15
6	-1.65	-1.56	-0.67	-0.95	-1.53	-1.66	-1.44	-1.64
7	-0.83	-1.03	-0.56	-0.77	-0.88	-0.78	-0.84	-0.64
8	1.06	1.23	0.57	0.93	1.08	1.01	1.06	0.88
9	-0.17	-0.23	-0.37	-0.43	-0.25	-0.16	-0.26	-0.10
10	-1.31	-1.40	-0.64	-0.90	-1.29	-1.28	-1.20	-1.17
11	-0.56	-0.74	-0.50	-0.66	-0.63	-0.52	-0.61	-0.39
12	0.66	0.85	0.13	0.37	0.63	0.62	0.60	0.48
13	1.43	1.47	1.17	1.57	1.52	1.41	1.51	1.33
14	-0.18	-0.25	-0.38	-0.44	-0.26	-0.17	-0.27	-0.11
15	-0.07	-0.10	-0.33	-0.36	-0.15	-0.06	-0.17	-0.04
16	-0.41	-0.55	-0.46	-0.58	-0.49	-0.38	-0.48	-0.27
17	0.51	0.68	0.01	0.19	0.46	0.47	0.44	0.35
18	0.97	1.16	0.46	0.80	0.98	0.92	0.96	0.78
19	-1.20	-1.33	-0.62	-0.88	-1.20	-1.16	-1.12	-1.54
20	-0.10	-0.14	-0.34	-0.38	-0.18	-0.09	-0.20	-0.06
21	1.53	1.51	1.37	1.76	1.64	1.52	1.64	1.46
22	-0.68	-0.87	-0.53	-0.71	-0.74	-0.64	-0.71	-0.49
23	1.70	1.58	1.77	2.11	1.85	1.71	1.85	1.71
24	1.40	1.45	1.11	1.52	1.48	1.38	1.48	1.29
25	0.17	0.23	-0.21	-0.16	0.10	0.16	0.07	0.10
26	-1.98	-1.65	-0.70	-0.98	-1.72	-2.04	-1.69	-2.15
27	-0.61	-0.79	-0.51	-0.68	-0.68	-0.57	-0.66	-0.43
28	1.03	1.21	0.53	0.89	1.05	0.98	1.03	0.84
29	0.66	0.85	0.13	0.37	0.63	0.62	0.60	0.48
30	-0.62	-0.81	-0.51	-0.69	-0.69	-0.58	-0.66	-0.44
31	-0.03	-0.04	-0.31	-0.33	-0.11	-0.03	-0.13	-0.02
32	0.15	0.21	-0.23	-0.18	0.08	0.14	0.05	0.09
33	-1.86	-1.62	-0.69	-0.97	-1.66	-1.90	-1.60	-1.96
34	-1.55	-1.52	-0.66	-0.94	-1.46	-1.54	-1.37	-1.49
35	0.34	0.46	-0.11	0.00	0.28	0.31	0.25	0.22
36	-0.63	-0.82	-0.52	-0.69	-0.69	-0.59	-0.67	-0.45
37	0.30	0.41	-0.14	-0.04	0.24	0.28	0.21	0.19
38	0.31	0.42	-0.13	-0.03	0.25	0.29	0.22	0.19
39	1.13	1.28	0.67	1.05	1.16	1.09	1.15	0.96
40	0.68	0.87	0.15	0.39	0.65	0.64	0.62	0.49
41	-0.01	-0.01	-0.30	-0.31	-0.09	-0.01	-0.11	-0.01
42	-0.45	-0.60	-0.47	-0.60	-0.52	-0.42	-0.52	-0.30
43	0.03	0.04	-0.29	-0.28	-0.05	0.03	-0.07	0.02
44	0.03	0.04	-0.29	-0.28	-0.05	0.03	-0.07	0.02
45	-0.64	-0.83	-0.52	-0.70	-0.70	-0.60	-0.68	-0.46
46	0.34	0.46	-0.11	0.00	0.28	0.31	0.25	0.22
47	0.01	0.01	-0.30	-0.30	0.01	0.01	0.01	0.01
48	0.49	0.65	-0.01	0.16	0.44	0.45	0.41	0.33
49	-0.63	-0.82	-0.52	-0.69	-0.69	-0.59	-0.67	-0.45
50	-2.07	-1.67	-0.70	-0.98	-1.76	-2.15	-1.76	-2.30
$\Sigma(x)$ $\Sigma(x^2)$	-3.55 44.3801	-2.21 47.3053	-6.18 16.9168	-5.19 30.6523	-4.02 42.6600	-3.84 43.4802	-3.90 40.4120	-4.95 40.8547

Tenth fifty

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	0.28	0.38	-0.15	-0.06	0.21	0.26	0.19	0.17
2	-1.52	-1.51	-0.66	-0.93	-1.44	-1.51	-1.35	-1.45
3	1.52	1.51	1.35	1.74	1.63	1.51	1.63	1.45
4	0.42	0.56	-0.06	0.09	0.36	0.39	0.34	0.28
5	-1.07	-1.24	-0.60	-0.85	-1.09	-1.03	-1.02	-0.89
6	-0.35	-0.47	-0.44	-0.55	-0.43	-0.32	-0.43	-0.23
7	-0.37	-0.50	-0.44	-0.56	-0.45	-0.34	-0.45	-0.24
8	-0.38	-0.51	-0.45	-0.57	-0.46	-0.35	-0.46	-0.25
9	0.09	0.12	-0.26	-0.23	0.02	0.08	-0.01	0.05
10	-0.38	-0.51	-0.45	-0.57	-0.46	-0.35	-0.46	0.25
11	0.38	0.51	-0.09	0.04	0.32	0.35	0.29	0.25
12	0.88	1.08	0.35	0.66	0.87	0.83	0.85	0.69
13	-1.92	-1.64	-0.69	-0.97	-1.69	-1.97	-1.64	-2.05
14	1.11	1.27	0.64	1.01	1.14	1.07	1.12	0.94
15	-1.59	-1.54	-0.67	-0.94	-1.49	-1.59	-1.40	-1.55
16	-0.33	-0.45	-0.43	-0.54	-0.41	-0.30	-0.41	-0.21
17	-0.63	-0.82	-0.52	-0.69	-0.69	-0.59	-0.67	-0.45
18	0.96	1.15	0.45	0.78	0.97	0.91	0.95	0.77
19	1.05	1.22	0.56	0.92	1.07	1.00	1.05	0.87
20	-0.48	-0.64	-0.48	-0.62	-0.55	-0.45	-0.54	-0.33
21	-0.60	-0.78	-0.51	-0.68	-0.67	-0.56	-0.65	-0.42
22	0.69	0.88	0.16	0.41	0.66	0.65	0.63	0.50
23	-0.46	-0.61	-0.47	-0.61	-0.53	-0.43	-0.53	-0.31
24	-0.87	-1.07	-0.57	-0.79	-0.92	-0.82	-0.87	-0.68
25	-0.45	-0.60	-0.47	-0.60	-0.52	-0.42	-0.52	-0.30
26	-0.14	-0.19	-0.36	-0.41	-0.22	-0.13	-0.23	-0.08
27	-1.02	-1.20	-0.60	-0.83	-1.05	-0.97	-0.99	-0.83
28	-0.58	-0.76	-0.50	-0.67	-0.65	-0.54	-0.63	-0.41
29	1.38	1.44	1.08	1.48	1.46	1.35	1.45	1.26
30	1.11	1.27	0.64	1.01	1.14	1.07	1.12	0.94
31	-1.49	-1.50	-0.66	-0.93	-1.42	-1.48	-1.33	-1.41
32	-0.14	-0.19	-0.36	-0.41	-0.22	-0.13	-0.23	-0.08
33	0.50	0.66	0.00	0.18	0.45	0.46	0.42	0.34
34	-0.35	-0.47	-0.44	-0.55	-0.43	-0.32	-0.43	-0.23
35	0.94	1.13	0.42	0.75	0.94	0.89	0.92	0.75
36	0.50	0.66	0.00	0.18	0.45	0.46	0.42	0.34
37	0.83	1.03	0.30	0.59	0.82	0.78	0.79	0.64
38	0.04	0.06	-0.28	-0.27	-0.04	0.04	-0.06	0.02
39	-0.12	-0.17	-0.35	-0.40	-0.20	-0.11	-0.21	-0.07
40	0.30	0.41	-0.14	-0.04	0.24	0.28	0.21	0.19
41	-1.45	-1.48	-0.65	-0.92	-1.39	-1.43	-1.30	-1.36
42	-0.23	-0.32	-0.40	-0.47	-0.31	-0.21	-0.32	-0.14
43	0.77	0.97	0.24	0.51	0.75	0.72	0.73	0.58
44	-0.37	-0.50	-0.44	-0.56	-0.45	-0.34	-0.45	-0.24
45	1.26	1.37	0.87	1.27	1.32	1.22	1.30	1.11
46	-0.62	-0.81	-0.51	-0.69	-0.69	-0.58	-0.66	-0.44
47	0.00	0.00	-0.30	-0.31	-0.08	0.00	-0.10	0.00
48	0.83	1.03	0.30	0.59	0.82	0.78	0.79	0.64
49	0.88	1.08	0.35	0.66	0.87	0.83	0.85	0.69
50	1.21	1.34	0.79	1.18	1.26	1.17	1.24	1.05
$\Sigma(x)$	0.02	0.65	-5.90	-4.17	-1.18	-0.17	-1.06	0.12
$\Sigma(x^2)$	36.5678	44.7135	13.5620	26.8849	36.8026	34.5703	34.4902	28.4536

Eleventh fifty

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	0.88	1.08	0.35	0.66	0.87	0.83	0.85	0.69
2	-0.62	-0.81	-0.51	-0.69	-0.69	-0.58	-0.66	-0.44
3	-0.36	-0.49	-0.44	-0.55	-0.44	-0.33	-0.44	-0.23
4	0.86	1.06	0.33	0.64	0.85	0.81	0.83	0.67
5	0.98	1.17	0.47	0.81	0.99	0.93	0.97	0.79
6	-0.73	-0.93	-0.54	-0.74	-0.79	-0.68	-0.76	-0.54
7	-0.52	-0.69	-0.49	-0.64	-0.59	-0.48	-0.58	-0.36
8	-0.72	-0.92	-0.54	-0.73	-0.78	-0.67	-0.75	-0.53
9	-0.97	-1.16	-0.59	-0.82	-1.00	-0.92	-0.95	-0.78
10	1.28	1.38	0.90	1.30	1.34	1.25	1.33	1.14
11	-0.18	-0.25	-0.38	-0.44	-0.26	-0.17	-0.27	-0.11
12	-0.27	-0.37	-0.41	-0.50	-0.35	-0.25	-0.36	-0.17
13	1.17	1.31	0.73	1.11	1.21	1.13	1.20	1.00
14	0.91	1.10	0.39	0.71	0.91	0.86	0.89	0.72
15	0.15	0.21	-0.23	-0.18	0.08	0.14	0.05	0.09
16	0.72	0.92	0.19	0.44	0.69	0.67	0.67	0.53
17	-1.25	-1.37	-0.63	-0.89	-1.24	-1.21	-1.16	-1.10
18	-1.08	-1.25	-0.61	-0.85	-1.10	-1.04	-1.03	-0.90
19	-1.18	-1.32	-0.62	-0.87	-1.18	-1.14	-1.11	-1.02
20	-0.89	-1.09	-0.57	-0.79	-0.93	-0.84	-0.88	-0.70
21	0.30	0.41	-0.14	-0.04	0.24	0.28	0.21	0.19
22	-0.65	-0.84	-0.52	-0.70	-0.71	-0.61	-0.69	-0.47
23	0.88	1.08	0.35	0.66	0.87	0.83	0.85	0.69
24	0.98	1.17	0.47	0.81	0.99	0.93	0.97	0.79
25	1.97	1.65	2.56	2.71	2.19	2.03	2.20	2.14
26	-0.82	-1.02	-0.56	-0.77	-0.87	-0.77	-0.83	-0.63
27	0.60	0.78	0.08	0.29	0.56	0.56	0.53	0.42
28	-0.37	-0.50	-0.44	-0.56	-0.45	-0.34	-0.45	-0.24
29	0.68	0.87	0.15	0.39	0.65	0.64	0.62	0.49
30	-0.88	-1.08	-0.57	-0.79	-0.92	-0.83	-0.88	-0.69
31	1.73	1.59	1.85	2.17	1.89	1.75	1.89	1.75
32	0.38	0.51	-0.09	0.04	0.32	0.35	0.29	0.25
33	-0.65	-0.84	-0.52	-0.70	-0.71	-0.61	-0.69	-0.47
34	1.07	1.24	0.59	0.95	1.09	1.03	1.08	0.89
35	-0.05	-0.07	-0.32	-0.35	-0.13	-0.05	-0.15	-0.03
36	-0.54	-0.71	-0.49	-0.65	-0.61	-0.50	-0.60	-0.37
37	-0.28	-0.38	-0.41	-0.51	-0.36	-0.26	-0.36	-0.17
38	-0.61	-0.79	-0.51	-0.68	-0.68	-0.57	-0.66	-0.43
39	-0.22	-0.30	-0.39	-0.47	-0.30	-0.20	-0.31	-0.14
40	0.15	0.21	-0.23	-0.18	0.08	0.14	0.05	0.09
41	-0.86	-1.06	-0.57	-0.78	-0.91	-0.81	-0.86	-0.67
42	0.75	0.95	0.22	0.48	0.73	0.70	0.70	0.56
43	0.82	1.02	0.29	0.58	0.81	0.77	0.78	0.63
44	-0.73	-0.93	-0.54	-0.74	-0.79	-0.68	-0.76	-0.54
45	0.40	0.54	-0.07	0.07	0.34	0.37	0.32	0.26
46	0.17	0.23	-0.21	-0.16	0.10	0.16	0.07	0.10
47	1.36	1.43	1.04	1.44	1.44	1.33	1.43	1.24
48	-0.98	-1.17	-0.59	-0.82	-1.01	-0.93	-0.95	-0.79
49	-1.30	-1.40	-0.64	-0.90	-1.28	-1.27	-1.19	-1.16
50	0.06	0.08	-0.27	-0.26	-0.02	0.06	-0.04	0.03
$\Sigma(x)$ $\Sigma(x^2)$	1.54 35.9004	0.25 46.4311	-3.68 21.2352	-2.49 35.1287	0.14 39.0552	1.81 33.5157	0.41 37.0161	2.47 26.6425

Twelfth fifty

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	-1.01	-1.19	-0.59	-0.83	-1.04	-0.96	-0.98	-0.82
2	-0.43	-0.58	-0.46	-0.59	-0.50	-0.40	-0.50	-0.29
3	-0.76	-0.96	-0.55	-0.75	-0.82	-0.71	-0.78	-0.57
4	0.13	0.18	-0.24	-0.20	0.06	0.12	0.03	0.08
5	-0.87	-1.07	-0.57	-0.79	-0.92	-0.82	-0.87	-0.68
6	1.69	1.57	1.74	2.09	1.84	1.70	1.84	1.69
7	0.02	0.03	-0.29	-0.29	-0.06	0.02	-0.08	0.01
8	0.33	0.45	-0.12	-0.01	0.27	0.30	0.24	0.21
9	-0.79	-0.99	-0.55	-0.76	-0.84	-0.74	-0.80	-0.60
10	1.41	1.46	1.13	1.53	1.50	1.39	1.49	1.30
11	1.79	1.60	2.01	2.30	1.96	1.82	1.97	1.85
12	1.38	1.44	1.08	1.48	1.46	1.35	1.45	1.26
13	0.71	0.90	0.18	0.43	0.68	0.66	0.66	0.52
14	-0.28	-0.38	-0.41	-0.51	-0.36	-0.26	-0.36	-0.17
15	-0.96	-1.15	-0.59	-0.82	-0.99	-0.91	-0.94	-0.77
16	-0.09	-0.12	-0.34	-0.38	-0.17	-0.08	-0.19	-0.05
17	-0.60	-0.78	-0.51	-0.68	-0.67	-0.56	-0.65	-0.42
18	-0.09	-0.12	-0.34	-0.38	-0.17	-0.08	-0.19	-0.05
19	-0.51	-0.68	-0.49	-0.64	-0.58	-0.47	-0.57	-0.35
20	-0.29	-0.40	-0.42	-0.51	-0.37	-0.27	-0.37	-0.18
21	1.50	1.50	1.31	1.71	1.61	1.49	1.60	1.42
22	-0.47	-0.63	-0.47	-0.62	-0.54	-0.44	-0.54	-0.32
23	-2.34	-1.70	-0.72	-0.99	-1.84	-2.48	-1.99	-2.80
24	-1.09	-1.25	-0.61	-0.85	-1.11	-1.05	-1.04	-0.91
25	1.18	1.32	0.74	1.13	1.22	1.14	1.21	1.02
26	-0.53	-0.70	-0.49	-0.65	-0.60	-0.49	-0.59	-0.37
27	-0.73	-0.93	-0.54	-0.74	-0.79	-0.68	-0.76	-0.54
28	1.47	1.49	1.25	1.65	1.57	1.45	1.56	1.38
29	-0.78	-0.98	-0.55	-0.75	-0.83	-0.73	-0.80	-0.59
30	-0.30	-0.41	-0.42	-0.52	-0.38	-0.28	-0.38	-0.19
31	1.73	1.59	1.85	2.17	1.89	1.75	1.89	1.75
32	1.02	1.20	0.52	0.87	1.03	0.97	1.02	0.83
33	-0.31	-0.42	-0.42	-0.52	-0.39	-0.29	-0.39	-0.19
34	1.28	1.38	0.90	1.30	1.34	1.25	1.33	1.14
35	-1.14	-1.29	-0.61	-0.86	-1.15	-1.10	-1.08	-0.97
36	-0.27	-0.37	-0.41	-0.50	-0.35	-0.25	-0.36	-0.17
37	-1.14	-1.29	-0.61	-0.86	-1.15	-1.10	-1.08	-0.97
38	1.68	1.57	1.72	2.07	1.83	1.69	1.83	1.68
39	-1.70	-1.58	-0.68	-0.95	-1.56	-1.71	-1.48	-1.71
40	0.27	0.37	-0.16	-0.07	0.20	0.25	0.18	0.17
41	0.69	0.88	0.16	0.41	0.66	0.65	0.63	0.50
42	-0.85	-1.05	-0.57	-0.78	-0.90	-0.80	-0.85	-0.66
43	0.50	0.66	0.00	0.18	0.45	0.46	0.42	0.34
44	-0.05	-0.07	-0.32	-0.35	-0.13	-0.05	-0.15	-0.03
45	-0.15	-0.21	-0.36	-0.42	-0.23	-0.14	-0.24	-0.09
46	1.66	1.56	1.67	2.03	1.80	1.67	1.80	1.65
47	0.76	0.96	0.23	0.50	0.74	0.71	0.71	0.57
48	-1.38	-1.44	-0.65	-0.91	-1.34	-1.35	-1.25	-1.26
49	0.53	0.70	0.02	0.21	0.48	0.49	0.46	0.37
50	-0.87	-1.07	-0.57	-0.79	-0.92	-0.82	-0.87	-0.68
$\Sigma(x)$	0.95	-1.00	0.88	1.79	0.89	1.31	1.19	2.34
$\Sigma(x^2)$	51.6413	55.3842	31.7714	51.7893	54.4177	50.5009	52.9947	46.1012

Thirteenth fifty

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	0.30	0.41	-0.14	-0.04	0.24	0.28	0.21	0.19
2	0.95	1.14	0.43	0.77	0.95	0.90	0.93	0.76
3	-1.00	-1.18	-0.59	-0.83	-1.03	-0.95	-0.97	-0.81
4	1.43	1.47	1.17	1.57	1.52	1.41	1.51	1.33
5	1.06	1.23	0.57	0.93	1.08	1.01	1.06	0.88
6	-0.66	-0.85	-0.52	-0.71	-0.72	-0.62	-0.70	-0.48
7	0.60	0.78	0.08	0.29	0.56	0.56	0.53	0.42
8	1.11	1.27	0.64	1.01	1.14	1.07	1.12	0.93
9	-0.08	-0.11	-0.34	-0.37	-0.16	-0.07	-0.18	-0.05
10	1.41	1.46	1.13	1.53	1.50	1.39	1.49	1.30
11	0.78	0.98	0.25	0.52	0.76	0.73	0.74	0.59
12	-2.01	-1.66	-0.70	-0.98	-1.73	-2.08	-1.71	-2.20
13	-1.53	-1.51	-0.66	-0.93	-1.45	-1.52	-1.36	-1.46
14	0.30	0.41	-0.14	-0.04	0.24	0.28	0.21	0.19
15	0.87	1.07	0.34	0.65	0.86	0.82	0.84	0.68
16	-0.77	-0.97	-0.55	-0.75	-0.82	-0.72	-0.79	-0.58
17	2.32	1.70	3.94	3.59	2.63	2.46	2.68	2.75
18	0.62	0.81	0.10	0.32	0.58	0.58	0.56	0.44
19	-0.93	-1.12	-0.58	-0.81	-0.97	-0.88	-0.92	-0.74
20	1.65	1.56	1.65	2.01	1.79	1.66	1.79	1.64
21	0.13	0.18	-0.24	-0.20	0.06	0.12	0.03	0.08
22	1.35	1.43	1.02	1.42	1.42	1.32	1.41	1.22
23	-1.14	-1.29	-0.61	-0.86	-1.15	-1.10	-1.08	-0.97
24	0.35	0.47	-0.11	0.01	0.29	0.32	0.26	0.23
25	-2.15	-1.68	-0.71	-0.98	-1.79	-2.25	-1.83	-2.44
26	0.43	0.58	-0.05	0.10	0.37	0.40	0.35	0.29
27	-1.41	-1.46	-0.65	-0.92	-1.36	-1.39	-1.27	-1.30
28	-0.14	-0.19	-0.36	-0.41	-0.22	-0.13	-0.23	-0.08
29	1.69	1.57	1.74	2.09	1.84	1.70	1.84	1.69
30	0.84	1.04	0.31	0.61	0.83	0.79	0.81	0.65
31	-2.14	-1.68	-0.71	-0.98	-1.79	-2.24	-1.82	-2.43
32	-0.75	-0.95	-0.54	-0.74	-0.81	-0.70	-0.77	-0.56
33	-0.58	-0.76	-0.50	-0.67	-0.65	-0.54	-0.63	-0.41
34	-0.25	-0.34	-0.40	-0.49	-0.33	-0.23	-0.34	-0.16
35	0.29	0.40	-0.14	-0.05	0.23	0.27	0.20	0.18
36	-1.00	-1.18	-0.59	-0.83	-1.03	-0.95	-0.97	-0.81
37	-1.69	-1.57	-0.68	-0.95	-1.56	-1.70	-1.47	-1.69
38	0.63	0.82	0.11	0.33	0.59	0.59	0.57	0.45
39	-1.34	-1.42	-0.64	-0.91	-1.31	-1.31	-1.22	-1.21
40	-1.76	-1.60	-0.68	-0.96	-1.60	-1.78	-1.52	-1.80
41	0.73	0.93	0.20	0.46	0.70	0.68	0.68	0.54
42	-0.10	-0.14	-0.34	-0.38	-0.18	-0.09	-0.20	-0.06
43	0.69	0.88	0.16	0.41	0.66	0.65	0.63	0.50
44	0.63	0.82	0.11	0.33	0.59	0.59	0.57	0.45
45	-1.33	-1.41	-0.64	-0.90	-1.30	-1.30	-1.21	-1.20
46	-0.63	-0.82	-0.52	-0.69	-0.69	-0.59	-0.67	-0.45
47	0.11	0.15	-0.25	-0.22	0.04	0.10	0.01	0.06
48	-1.08	-1.25	-0.61	-0.85	-1.10	-1.04	-1.03	-0.90
49	1.05	1.22	0.56	0.92	1.07	1.00	1.05	0.87
50	0.48	0.64	-0.02	0.15	0.43	0.45	0.40	0.33
$\Sigma(x)$ $\Sigma(x^2)$	-1.67 61.4693	0.28 62.0998	0.30 34.5500	1.57 48.1979	-0.78 59.7628	-2.05 61.2141	-0.41 57.3189	-3.15 58.3595

Fourteenth fifty

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	0.84	1.04	0.31	0.61	0.83	0.79	0.81	0.65
2	0.77	0.97	0.24	0.51	0.75	0.72	0.73	0.58
3	0.78	0.98	0.25	0.52	0.76	0.73	0.74	0.59
4	1.78	1.60	1.98	2.28	1.95	1.81	1.96	1.83
5	-0.40	-0.54	-0.45	-0.58	-0.48	-0.37	-0.47	-0.26
6	0.08	0.11	-0.26	-0.24	0.01	0.07	-0.02	0.05
7	-0.01	-0.01	-0.30	-0.31	-0.09	-0.01	-0.11	-0.01
8	-0.14	-0.19	-0.36	-0.41	-0.22	-0.13	-0.23	-0.08
9	0.11	0.15	-0.25	-0.22	0.04	0.10	0.01	0.06
10	0.19	0.26	-0.20	-0.14	0.12	0.17	0.09	0.12
11	-0.09	-0.12	-0.34	-0.38	-0.17	-0.08	-0.19	-0.05
12	-0.40	-0.54	-0.45	-0.58	-0.48	-0.37	-0.47	-0.26
13	0.46	0.61	-0.03	0.13	0.41	0.43	0.38	0.31
14	0.07	0.10	-0.27	-0.25	0.00	0.06	-0.03	0.04
15	-1.95	-1.64	-0.70	-0.97	-1.70	-2.01	-1.67	-2.10
16	-1.09	-1.25	-0.61	-0.85	-1.11	-1.05	-1.04	-0.91
17	1.54	1.52	1.40	1.79	1.65	1.53	1.65	1.48
18	0.45	0.60	-0.04	0.12	0.40	0.42	0.37	0.30
19	-0.61	-0.79	-0.51	-0.68	-0.68	-0.57	-0.66	-0.43
20	-0.53	-0.70	-0.49	-0.65	-0.60	-0.49	-0.59	-0.37
21	0.54	0.71	0.03	0.22	0.49	0.50	0.47	0.37
22	1.49	1.50	1.29	1.69	1.59	1.48	1.59	1.41
23	0.53	0.70	0.02	0.21	0.48	0.49	0.46	0.37
24	2.40	1.70	4.34	3.80	2.73	2.55	2.79	2.91
25	-0.99	-1.17	-0.59	-0.82	-1.02	-0.94	-0.96	-0.80
26	0.79	0.99	0.26	0.54	0.77	0.74	0.75	0.60
27	0.07	0.10	-0.27	-0.25	0.00	0.06	-0.03	0.04
28	0.82	1.02	0.29	0.58	0.81	0.77	0.78	0.63
29	0.69	0.88	0.16	0.41	0.66	0.65	0.63	0.50
30	-0.44	-0.59	-0.46	-0.60	-0.51	-0.41	-0.51	-0.29
31	1.20	1.33	0.77	1.16	1.25	1.16	1.23	1.04
32	1.03	1.21	0.53	0.89	1.05	0.98	1.03	0.84
33	-1.25	-1.37	-0.63	-0.89	-1.24	-1.21	-1.16	-1.10
34	-0.89	-1.09	-0.57	-0.79	-0.93	-0.84	-0.88	-0.70
35	0.78	0.98	0.25	0.52	0.76	0.73	0.74	0.59
36	2.04	1.66	2.79	2.88	2.28	2.11	2.30	2.25
37	0.81	1.01	0.28	0.57	0.79	0.76	0.77	0.62
38	-0.93	-1.12	-0.58	-0.81	-0.97	-0.88	-0.92	-0.74
39	0.53	0.70	0.02	0.21	0.48	0.49	0.46	0.37
40	2.23	1.69	3.54	3.35	2.52	2.35	2.55	2.59
41	1.27	1.38	0.88	1.28	1.33	1.23	1.32	1.12
42	0.28	0.38	-0.15	-0.06	0.21	0.26	0.19	0.18
43	-1.38	-1.44	-0.65	-0.91	-1.34	-1.35	-1.25	-1.26
44	-0.77	-0.97	-0.55	-0.75	-0.82	-0.72	-0.79	-0.58
45	0.93	1.12	0.41	0.74	0.93	0.88	0.91	0.74
46	2.90	1.73	7.64	5.27	3.32	3.15	3.55	3.94
47	0.62	0.81	0.10	0.32	0.58	0.58	0.56	0.44
48	-1.30	-1.40	-0.64	-0.90	-1.28	-1.27	-1.19	-1.16
49	-0.08	-0.11	-0.34	-0.37	-0.16	-0.07	-0.18	-0.05
50	0.04	0.06	-0.28	-0.27	-0.04	0.04	-0.06	0.02
$\Sigma(x)$	15.81	14.56	16.81	16.92	16.11	16.02	16.41	16.43
$\Sigma(x^2)$	57.3619	52.9306	112.8177	89.2754	64.7851	58.6566	65.1299	62.5201

Fifteenth fifty

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	-0.01	-0.01	-0.30	-0.31	-0.09	-0.01	-0.11	-0.01
2	0.08	0.11	-0.26	-0.24	0.01	0.07	-0.02	0.05
3	2.51	1.71	4.93	4.12	2.86	2.69	2.95	3.13
4	-0.88	-1.08	-0.57	-0.79	-0.92	-0.83	-0.88	-0.69
5	0.64	0.83	0.11	0.34	0.60	0.60	0.58	0.46
6	0.14	0.19	-0.23	-0.19	0.07	0.13	0.04	0.08
7	0.82	1.02	0.29	0.58	0.81	0.77	0.78	0.63
8	-0.51	-0.68	-0.49	-0.64	-0.58	-0.47	-0.57	-0.35
9	-1.75	-1.59	-0.68	-0.96	-1.59	-1.77	-1.52	-1.78
10	0.47	0.63	-0.02	0.14	0.42	0.44	0.39	0.32
11	1.13	1.28	0.67	1.05	1.16	1.09	1.15	0.96
12	1.13	1.28	0.67	1.05	1.16	1.09	1.15	0.96
13	0.40	0.54	-0.07	0.07	0.34	0.37	0.32	0.26
14	-1.77	-1.60	-0.68	-0.96	-1.61	-1.79	-1.53	-1.81
15	-0.06	-0.08	-0.33	-0.35	-0.14	-0.06	-0.16	-0.03
16	0.47	0.63	-0.02	0.14	0.42	0.44	0.39	0.32
17	-0.18	-0.25	-0.38	-0.44	-0.26	-0.17	-0.27	-0.11
18	0.30	0.41	-0.14	-0.04	0.24	0.28	0.21	0.19
19	0.14	0.19	-0.23	-0.19	0.07	0.13	0.04	0.08
20	0.54	0.71	0.03	0.22	0.49	0.50	0.47	0.37
21	-0.07	-0.10	-0.33	-0.36	-0.15	-0.06	-0.17	-0.04
22	0.17	0.23	-0.21	-0.16	0.10	0.16	0.07	0.10
23	1.07	1.24	0.59	0.95	1.09	1.03	1.08	0.89
24	2.75	1.72	6.47	4.81	3.15	2.98	3.31	3.62
25	-2.67	-1.72	-0.73	-1.00	-1.86*	-2.88	-2.33	-3.45
26	-0.43	-0.58	-0.46	-0.59	-0.50	-0.40	-0.50	-0.29
27	-0.03	-0.04	-0.31	-0.33	-0.11	-0.03	-0.13	-0.02
28	-1.10	-1.26	-0.61	-0.85	-1.12	-1.06	-1.05	-0.92
29	-0.24	-0.33	-0.40	-0.48	-0.32	-0.22	-0.33	-0.15
30	0.47	0.63	-0.02	0.14	0.42	0.44	0.39	0.32
31	1.14	1.29	0.68	1.06	1.17	1.10	1.16	0.97
32	-0.64	-0.83	-0.52	-0.70	-0.70	-0.60	-0.68	-0.46
33	-0.46	-0.61	-0.47	-0.61	-0.53	-0.43	-0.53	-0.31
34	-0.12	-0.17	-0.35	-0.40	-0.20	-0.11	-0.21	-0.07
35	1.33	1.41	0.99	1.39	1.40	1.30	1.39	1.20
36	-1.05	-1.22	-0.60	-0.84	-1.07	-1.00	-1.01	-0.87
37	-1.20	-1.33	-0.62	-0.88	-1.20	-1.16	-1.12	-1.04
38	0.75	0.95	0.22	0.48	0.73	0.70	0.70	0.56
39	0.11	0.15	-0.25	-0.22	0.04	0.10	0.01	0.06
40	0.44	0.59	-0.04	0.11	0.39	0.41	0.36	0.29
41	-2.39	-1.70	-0.72	-0.99	-1.85	-2.54	-2.04	-2.89
42	0.95	1.14	0.43	0.77	0.95	0.90	0.93	0.76
43	-0.54	-0.71	-0.49	-0.65	-0.61	-0.50	-0.60	-0.37
44	-0.93	-1.12	-0.58	-0.81	-0.97	-0.88	-0.92	-0.74
45	-0.86	-1.06	-0.57	-0.78	-0.91	-0.81	-0.86	-0.67
46	-0.97	-1.16	-0.59	-0.82	-1.00	-0.92	-0.95	-0.78
47	-1.28	-1.38	-0.63	-0.89	-1.26	-1.25	-1.18	-1.14
48	0.08	0.11	-0.26	-0.24	0.01	0.07	-0.02	0.05
49	-0.12	-0.17	-0.35	-0.40	-0.20	-0.11	-0.21	-0.07
50	-0.50	-0.66	-0.48	-0.63	-0.57	-0.46	-0.56	-0.34
$\Sigma(x)$ $\Sigma(x^2)$	-2.73 54.1509	-2.45 46.5775	1.09 76.7501	-1.32 60.9554	-2.22 52.2892	-2.73 56.5775	-2.59 54.6291	-2.77 63.6593

Sixteenth fifty

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	0.24	0.33	-0.17	-0.10	0.17	0.22	0.15	0.15
2	-0.57	-0.75	-0.50	-0.67	-0.64	-0.53	-0.62	-0.40
3	-0.72	-0.92	-0.54	-0.73	-0.78	-0.67	-0.75	-0.53
4	-0.30	-0.41	-0.42	-0.52	-0.38	-0.28	-0.38	-0.19
5	0.88	1.08	0.35	0.66	0.87	0.83	0.85	0.69
6	0.30	0.41	-0.14	-0.04	0.24	0.28	0.21	0.19
7	1.28	1.38	0.90	1.30	1.34	1.25	1.33	1.14
8	-1.38	-1.44	-0.65	-0.91	-1.34	-1.35	-1.25	-1.26
9	1.68	1.57	1.72	2.07	1.83	1.69	1.83	1.68
10	-0.17	-0.23	-0.37	-0.43	-0.25	-0.16	-0.26	-0.10
11	-0.14	-0.19	-0.36	-0.41	-0.22	-0.13	-0.23	-0.08
12	0.68	0.87	0.15	0.39	0.65	0.64	0.62	0.50
13	0.18	0.25	-0.21	-0.15	0.11	0.17	0.08	0.11
14	-1.71	-1.58	-0.68	-0.96	-1.57	-1.72	-1.49	-1.73
15	-0.28	-0.38	-0.41	-0.51	-0.36	-0.26	-0.36	-0.18
16	0.94	1.13	0.42	0.75	0.94	0.89	0.92	0.75
17	0.33	0.45	-0.12	-0.01	0.27	0.30	0.24	0.21
18	1.18	1.32	0.74	1.13	1.22	1.14	1.21	1.02
19	-0.62	-0.81	-0.51	-0.69	-0.69	-0.58	-0.66	-0.44
20	1.69	1.57	1.74	2.09	1.84	1.70	1.84	1.69
21	-0.40	-0.54	-0.45	-0.58	-0.48	-0.37	-0.47	-0.26
22	-1.96	-1.65	-0.70	-0.97	-1.71	-2.02	-1.67	-2.12
23	-0.84	-1.04	-0.56	-0.78	-0.89	-0.79	-0.84	-0.65
24	-0.54	-0.71	-0.49	-0.65	-0.61	-0.50	-0.60	-0.37
25	0.05	0.07	-0.28	-0.27	-0.03	0.05	-0.05	0.03
26	-0.82	-1.02	-0.56	-0.77	-0.87	-0.77	-0.83	-0.63
27	0.46	0.61	-0.03	0.13	0.41	0.43	0.38	0.31
28	-0.41	-0.55	-0.46	-0.58	-0.49	-0.38	-0.48	-0.27
29	1.27	1.38	0.88	1.28	1.33	1.23	1.32	1.12
30	-1.22	-1.35	-0.63	-0.88	-1.21	-1.18	-1.13	-1.06
31	1.76	1.60	1.93	2.24	1.93	1.78	1.93	1.80
32	-0.20	-0.27	-0.38	-0.45	-0.28	-0.18	-0.29	-0.12
33	1.05	1.22	0.56	0.92	1.07	1.00	1.05	0.87
34	-0.95	-1.14	-0.58	-0.81	-0.99	-0.90	-0.93	-0.76
35	-0.16	-0.22	-0.37	-0.43	-0.24	-0.15	-0.25	-0.10
36	-1.11	-1.27	-0.61	-0.86	-1.12	-1.07	-1.05	-0.93
37	0.97	1.16	0.46	0.80	0.98	0.92	0.96	0.78
38	1.53	1.51	1.37	1.76	1.64	1.52	1.64	1.46
39	-0.05	-0.07	-0.32	-0.35	-0.13	-0.05	-0.15	-0.03
40	0.64	0.83	0.11	0.34	0.60	0.60	0.58	0.46
41	-1.83	-1.62	-0.69	-0.97	-1.64	-1.86	-1.58	-1.91
42	-0.73	-0.93	-0.54	-0.74	-0.79	-0.68	-0.76	-0.54
43	1.26	1.37	0.87	1.27	1.32	1.22	1.30	1.11
44	-0.11	-0.15	-0.35	-0.39	-0.19	-0.10	-0.20	-0.06
45	0.02	0.03	-0.29	-0.29	-0.06	0.02	-0.08	0.01
46	-1.52	-1.51	-0.66	-0.93	-1.44	-1.51	-1.35	-1.45
47	-1.61	-1.55	-0.67	-0.94	-1.51	-1.61	-1.42	-1.58
48	-0.83	-1.03	-0.56	-0.77	-0.88	-0.78	-0.84	-0.64
49	1.78	1.60	1.98	2.28	1.95	1.81	1.96	1.83
50	0.29	0.40	-0.14	-0.05	0.23	0.27	0.20	0.18
$\Sigma(x)$	-0.72	-1.19	-1.22	-0.18	-0.85	-0.62	-0.37	-0.30
$\Sigma(x^2)$	50.9664	54.6567	27.1922	45.1382	52.4567	49.6494	49.9199	44.7696

Seventeenth fifty

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	-0.75	-0.95	-0.54	-0.74	-0.81	-0.70	-0.77	-0.56
2	-0.47	-0.63	-0.47	-0.62	-0.54	-0.44	-0.54	-0.32
3	0.55	0.72	0.04	0.23	0.51	0.51	0.48	0.38
4	0.87	1.07	0.34	0.65	0.86	0.82	0.84	0.68
5	-0.63	-0.82	-0.52	-0.69	-0.69	-0.59	-0.67	-0.45
6	-0.60	-0.78	-0.51	-0.68	-0.67	-0.56	-0.65	-0.42
7	0.85	1.05	0.32	0.62	0.84	0.80	0.82	0.66
8	-0.86	-1.06	-0.57	-0.78	-0.91	-0.81	-0.86	-0.67
9	0.79	0.99	0.26	0.54	0.77	0.74	0.75	0.60
10	0.77	0.97	0.24	0.51	0.75	0.72	0.73	0.58
11	-1.31	-1.40	-0.64	-0.90	-1.29	-1.28	-1.20	-1.17
12	1.02	1.20	0.52	0.87	1.03	0.97	1.02	0.83
13	1.01	1.19	0.51	0.86	1.02	0.96	1.00	0.82
14	-0.69	-0.88	-0.53	-0.72	-0.75	-0.65	-0.72	-0.50
15	-0.59	-0.77	-0.51	-0.67	-0.66	-0.55	-0.64	-0.42
16	-1.76	-1.60	-0.68	-0.96	-1.60	-1.78	-1.52	-1.80
17	0.47	0.63	-0.02	0.14	0.42	0.44	0.39	0.32
18	2.36	1.70	4.14	3.70	2.68	2.50	2.74	2.83
19	-0.88	-1.08	-0.57	-0.79	-0.92	-0.83	-0.88	-0.69
20	-0.42	-0.56	-0.46	-0.59	-0.50	-0.39	-0.49	-0.28
21	0.06	0.08	-0.27	-0.26	-0.02	0.06	-0.04	0.03
22	-1.55	-1.52	-0.66	-0.94	-1.46	-1.54	-1.37	-1.49
23	0.92	1.11	0.40	0.72	0.92	0.87	0.90	0.73
24	0.48	0.64	-0.02	0.15	0.43	0.45	0.40	0.33
25	0.88	1.08	0.35	0.66	0.87	0.83	0.85	0.69
26	-2.01	-1.66	-0.70	-0.98	-1.73	-2.08	-1.71	-2.20
27	1.62	1.55	1.57	1.95	1.75	1.62	1.75	1.59
28	0.40	0.54	-0.07	0.07	0.34	0.37	0.32	0.26
29	1.94	1.64	2.46	2.64	2.15	1.99	2.16	2.09
30	0.49	0.65	-0.01	0.16	0.44	0.45	0.41	0.33
31	0.36	0.49	-0.10	0.02	0.30	0.33	0.27	0.23
32	-0.47	-0.63	-0.47	-0.62	-0.54	-0.44	-0.54	-0.32
33	2.19	1.68	3.37	3.25	2.47	2.30	2.50	2.51
34	1.22	1.35	0.80	1.20	1.27	1.18	1.26	1.06
35	0.69	0.88	0.16	0.41	0.66	0.65	0.63	0.50
36	-0.70	-0.89	-0.53	-0.72	-0.76	-0.66	-0.73	-0.51
37	-1.27	-1.38	-0.63	-0.89	-1.25	-1.23	-1.17	-1.12
38	0.18	0.25	-0.21	0.15	0.11	0.17	0.08	0.11
39	0.98	1.17	0.47	0.81	0.99	0.93	0.97	0.79
40	-0.76	-0.96	-0.55	-0.75	-0.82	-0.71	-0.78	-0.57
41	-0.48	-0.64	-0.48	-0.62	-0.55	-0.45	-0.54	-0.33
42	1.49	1.50	1.29	1.69	1.59	1.48	1.59	1.41
43	-0.34	-0.46	-0.43	-0.54	-0.42	-0.31	-0.42	-0.22
44	0.15	0.21	-0.23	-0.18	0.08	0.14	0.05	0.09
45	-0.77	-0.97	-0.55	-0.75	-0.82	-0.72	-0.79	-0.58
46	1.28	1.38	0.90	1.30	1.34	1.25	1.33	1.14
47	-1.16	-1.31	-0.62	-0.87	-1.16	-1.12	-1.09	-0.99
48	1.88	1.63	2.27	2.50	2.08	1.92	2.09	1.99
49	-0.67	-0.86	-0.53	-0.71	-0.73	-0.63	-0.71	-0.49
50	-0.78	-0.98	-0.55	-0.75	-0.83	-0.73	-0.80	-0.59
$\Sigma(x)$	5.98	4.56	6.78	7.78	6.24	6.25	6.70	6.89
$\Sigma(x^2)$	56.3456	58.5270	53.9998	65.4558	60.9584	55.8065	59.2344	52.9191

Eighteenth fifty

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	-0.97	-1.16	-0.59	-0.82	-1.00	-0.92	-0.95	-0.78
2	2.38	1.70	4.24	3.74	2.70	2.53	2.76	2.86
3	0.78	0.98	0.25	0.52	0.76	0.73	0.74	0.59
4	0.35	0.47	-0.11	0.01	0.29	0.32	0.26	0.23
5	1.10	1.26	0.63	1.00	1.13	1.06	1.11	0.92
6	-0.03	-0.04	-0.31	-0.33	-0.11	-0.03	-0.13	-0.02
7	-1.47	-1.49	-0.66	-0.93	-1.41	-1.45	-1.32	-1.38
8	0.68	0.87	0.15	0.39	0.65	0.64	0.62	0.50
9	-0.73	-0.93	-0.54	-0.74	-0.79	-0.68	-0.76	-0.54
10	-1.17	-1.31	-0.62	-0.87	-1.17	-1.13	-1.10	-1.00
11	-0.05	-0.07	-0.32	-0.35	-0.13	-0.05	-0.15	-0.03
12	1.02	1.20	0.52	0.87	1.03	0.97	1.02	0.83
13	-0.19	-0.26	-0.38	-0.45	-0.27	-0.17	-0.28	-0.12
14	-0.89	-1.09	-0.57	-0.79	-0.93	-0.84	-0.88	-0.70
15	0.56	0.74	0.05	0.25	0.52	0.52	0.49	0.39
16	0.47	0.63	-0.02	0.14	0.42	0.44	0.39	0.32
17	-0.70	-0.89	-0.53	-0.72	-0.76	-0.66	-0.73	-0.51
18	1.49	1.50	1.29	1.69	1.59	1.48	1.59	1.41
19	-0.75	-0.95	-0.54	-0.74	-0.81	-0.70	-0.77	-0.56
20	1.03	1.21	0.53	0.89	1.05	0.98	1.03	0.84
21	0.24	0.33	-0.17	-0.10	0.17	0.22	0.15	0.15
22	-0.42	-0.56	-0.46	-0.59	-0.50	-0.39	-0.49	-0.28
23	-0.67	-0.86	-0.53	-0.71	-0.73	-0.63	-0.71	-0.49
24	-1.32	-1.41	-0.64	-0.90	-1.29	-1.29	-1.21	-1.19
25	0.20	0.27	-0.20	-0.13	0.13	0.18	0.10	0.12
26	1.31	1.40	0.95	1.35	1.38	1.28	1.37	1.17
27	1.11	1.27	0.64	1.01	1.14	1.07	1.12	0.93
28	-0.21	-0.29	-0.39	-0.46	-0.29	-0.19	-0.30	-0.13
29	-0.35	-0.47	-0.44	-0.55	-0.43	-0.32	-0.43	-0.23
30	-2.76	-1.72	-0.73	-1.00	-1.86*	-2.99	-2.44	-3.64
31	-0.63	-0.82	-0.52	-0.69	-0.69	-0.59	-0.67	-0.45
32	-2.31	-1.70	-0.72	-0.99	-1.83	-2.44	-1.96	-2.74
33	-0.01	-0.01	-0.30	-0.31	-0.09	-0.01	-0.11	-0.01
34	0.41	0.55	-0.07	0.08	0.35	0.38	0.33	0.27
35	3.02	1.73	8.72	5.65	3.45	3.29	3.74	4.21
36	1.26	1.37	0.87	1.27	1.32	1.22	1.30	1.11
37	0.36	0.49	-0.10	0.02	0.30	0.33	0.27	0.23
38	-0.43	-0.58	-0.46	-0.59	-0.50	-0.40	-0.50	-0.29
39	-0.26	-0.36	-0.41	-0.49	-0.34	-0.24	-0.35	-0.16
40	1.73	1.59	1.85	2.17	1.89	1.75	1.89	1.75
41	-0.64	-0.83	-0.52	-0.70	-0.70	-0.60	-0.68	-0.46
42	0.20	0.27	-0.20	-0.13	0.13	0.18	0.10	0.12
43	-2.08	-1.67	-0.71	-0.98	-1.76	-2.16	-1.77	-2.32
44	-0.46	-0.61	-0.47	-0.61	-0.53	-0.43	-0.53	-0.31
45	-0.84	-1.04	-0.56	-0.78	-0.89	-0.79	-0.84	-0.65
46	0.16	0.22	-0.22	-0.17	0.09	0.15	0.06	0.10
47	-1.42	-1.46	-0.65	-0.92	-1.37	-1.40	-1.28	-1.32
48	0.20	0.27	-0.20	-0.13	0.13	0.18	0.10	0.12
49	1.34	1.42	1.00	1.41	1.41	1.31	1.40	1.21
50	-0.63	-0.82	-0.52	-0.69	-0.69	-0.59	-0.67	-0.45
$\Sigma(x)$	-0.99	-1.66	6.31	3.10	0.16	-0.88	-0.07	-0.38
$\Sigma(x^2)$	62.9609	53.5924	111.1731	77.1048	62.0682	65.7264	65.6813	75.0474

Nineteenth fifty

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	-0.19	-0.26	-0.38	-0.45	-0.27	-0.17	-0.28	-0.12
2	0.13	0.18	-0.24	-0.20	0.06	0.12	0.03	0.08
3	-0.82	-1.02	-0.56	-0.77	-0.87	-0.77	-0.83	-0.63
4	-0.18	-0.25	-0.38	-0.44	-0.26	-0.17	-0.27	-0.11
5	-2.43	-1.71	-0.72	-0.99	-1.85	-2.59	-2.08	-2.97
6	0.78	0.98	0.25	0.52	0.76	0.73	0.74	0.59
7	-0.98	-1.17	-0.59	-0.82	-1.01	-0.93	-0.95	-0.79
8	-0.09	-0.12	-0.34	-0.38	-0.17	-0.08	-0.19	-0.05
9	-0.81	-1.01	-0.56	-0.77	-0.86	-0.76	-0.82	-0.62
10	0.56	0.74	0.05	0.25	0.52	0.52	0.49	0.39
11	-2.32	-1.70	-0.72	-0.99	-1.84	-2.46	-1.97	-2.75
12	0.01	0.01	-0.30	-0.30	-0.07	0.01	-0.09	0.01
13	-0.90	-1.09	-0.57	-0.80	-0.94	-0.85	-0.89	-0.71
14	0.22	0.30	-0.19	-0.12	0.15	0.20	0.13	0.14
15	-1.36	-1.43	-0.64	-0.91	-1.32	-1.33	-1.24	-1.24
16	0.65	0.84	0.12	0.36	0.62	0.61	0.59	0.47
17	0.39	0.53	-0.08	0.05	0.33	0.36	0.30	0.26
18	0.53	0.70	0.02	0.21	0.48	0.49	0.46	0.37
19	0.81	1.01	0.28	0.57	0.79	0.76	0.77	0.62
20	0.91	1.10	0.39	0.71	0.91	0.86	0.89	0.72
21	-0.14	-0.19	-0.36	-0.41	-0.22	-0.13	-0.23	-0.08
22	1.76	1.60	1.93	2.24	1.93	1.78	1.93	1.80
23	1.19	1.33	0.76	1.15	1.23	1.15	1.22	1.03
24	1.41	1.46	1.13	1.53	1.50	1.39	1.49	1.30
25	-1.00	-1.18	-0.59	-0.83	-1.03	-0.95	-0.97	-0.81
26	-1.20	-1.33	-0.62	-0.88	-1.20	-1.16	-1.12	-1.04
27	0.71	0.90	0.18	0.43	0.68	0.66	0.66	0.52
28	-0.59	-0.77	-0.51	-0.67	-0.66	-0.55	-0.64	-0.42
29	-1.38	-1.44	-0.65	-0.91	-1.34	-1.35	-1.25	-1.26
30	2.17	1.68	3.29	3.20	2.44	2.27	2.47	2.48
31	0.30	0.41	-0.14	-0.04	0.24	0.28	0.21	0.19
32	0.51	0.68	0.01	0.19	0.46	0.47	0.44	0.35
33	-0.07	-0.10	-0.33	-0.36	-0.15	-0.06	-0.17	-0.04
34	0.55	0.72	0.04	0.23	0.51	0.51	0.48	0.38
35	0.52	0.69	0.02	0.20	0.47	0.48	0.45	0.36
36	0.01	0.01	-0.30	-0.30	-0.07	0.01	-0.09	0.01
37	1.37	1.44	1.06	1.46	1.45	1.34	1.44	1.25
38	0.82	1.02	0.29	0.58	0.81	0.77	0.78	0.63
39	-0.08	-0.11	-0.34	-0.37	-0.16	-0.07	-0.18	-0.05
40	-0.25	-0.34	-0.40	-0.49	-0.33	-0.23	-0.34	-0.16
41	1.32	1.41	0.97	1.37	1.39	1.29	1.38	1.19
42	-0.50	-0.66	-0.48	-0.63	-0.57	-0.46	-0.56	-0.34
43	-0.04	-0.06	-0.32	-0.34	-0.12	-0.04	-0.14	-0.02
44	0.98	1.17	0.47	0.81	0.99	0.93	0.97	0.79
45	-0.50	-0.66	-0.48	-0.63	-0.57	0.46	-0.56	-0.34
46	0.51	0.68	0.01	0.19	0.46	0.47	0.44	0.35
47	1.59	1.54	1.51	1.88	1.72	1.59	1.71	1.55
48	0.44	0.59	-0.04	0.11	0.39	0.41	0.36	0.29
49	0.97	1.16	0.46	0.80	0.98	0.92	0.96	0.78
50	-1.29	-1.39	-0.64	-0.90	-1.27	-1.26	-1.19	-1.15
$\Sigma(x)$ $\Sigma(x^2)$	5.00 48.4038	6.89 49.4833	0.77 27.9639	3.34 40.9940	5.12 47.0018	5.47 48.2137	4.74 46.6414	3.20 46.8674

Twentieth fifty

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	1.31	1.40	0.95	1.35	1.38	1.28	1.37	1.17
2	-0.17	-0.23	-0.37	-0.43	-0.25	-0.16	-0.26	-0.10
3	-0.95	-1.14	-0.58	-0.81	-0.99	-0.90	-0.93	-0.76
4	0.69	0.88	0.16	0.41	0.66	0.65	0.63	0.50
5	0.32	0.43	-0.13	-0.02	0.26	0.30	0.23	0.20
6	0.26	0.36	-0.16	-0.08	0.19	0.24	0.17	0.16
7	-0.39	-0.53	-0.45	-0.57	-0.47	-0.36	-0.46	-0.26
8	-0.50	-0.66	-0.48	-0.63	-0.57	-0.46	-0.56	-0.34
9	1.91	1.63	2.36	2.57	2.11	1.96	2.13	2.04
10	2.37	1.70	4.19	3.72	2.69	2.52	2.75	2.85
11	1.77	1.60	1.95	2.26	1.94	1.79	1.94	1.81
12	-0.09	-0.12	-0.34	-0.38	-0.17	-0.08	-0.19	-0.05
13	-0.62	-0.81	-0.51	-0.69	-0.69	-0.58	-0.66	-0.44
14	-1.02	-1.20	-0.60	-0.83	-1.05	-0.97	-0.99	-0.83
15	-0.54	-0.71	-0.49	-0.65	-0.61	-0.50	-0.60	-0.37
16	0.07	0.10	-0.27	-0.25	0.00	0.06	-0.03	0.04
17	-0.16	-0.22	-0.37	-0.43	-0.24	-0.15	-0.25	-0.10
18	-1.44	-1.47	-0.65	-0.92	-1.39	-1.42	-1.29	-1.34
19	-0.60	-0.78	-0.51	-0.68	-0.67	-0.56	-0.65	-0.42
20	-0.23	-0.32	-0.40	-0.47	-0.31	-0.21	-0.32	-0.14
21	-0.67	-0.86	-0.53	-0.71	-0.73	-0.63	-0.71	-0.49
22	1.56	1.53	1.44	1.82	1.68	1.55	1.68	1.51
23	0.51	0.68	0.01	0.19	0.46	0.47	0.44	0.35
24	0.33	0.45	-0.12	-0.01	0.27	0.30	0.24	0.21
25	-0.01	-0.01	-0.30	-0.31	-0.09	-0.01	-0.11	-0.01
26	0.13	0.18	-0.24	-0.20	0.06	0.12	0.03	0.08
27	0.81	1.01	0.28	0.57	0.79	0.76	0.77	0.62
28	-1.09	-1.25	-0.61	-0.85	-1.11	-1.05	-1.04	-0.91
29	-0.90	-1.09	-0.57	-0.80	-0.94	-0.85	-0.89	-0.71
30	-0.39	-0.53	-0.45	-0.57	-0.47	-0.36	-0.46	-0.26
31	0.57	0.75	0.06	0.26	0.53	0.53	0.50	0.40
32	-0.88	-1.08	-0.57	-0.79	-0.92	-0.83	-0.88	-0.69
33	0.57	0.75	0.06	0.26	0.53	0.53	0.50	0.40
34	0.98	1.17	0.47	0.81	0.99	0.93	0.97	0.79
35	-0.38	-0.51	-0.45	-0.57	-0.46	-0.35	-0.46	-0.25
36	-1.06	-1.23	-0.60	-0.84	-1.08	-1.01	-1.02	-0.88
37	1.79	1.60	2.01	2.30	1.96	1.82	1.97	1.85
38	0.00	0.00	-0.30	-0.31	-0.08	0.00	-0.10	0.00
39	0.23	0.32	-0.18	-0.11	0.16	0.21	0.14	0.14
40	0.21	0.29	-0.19	-0.12	0.14	0.19	0.11	0.13
41	-0.76	-0.96	-0.55	-0.75	-0.82	-0.71	-0.78	-0.57
42	-0.94	-1.13	-0.58	-0.81	-0.98	-0.89	-0.92	-0.75
43	-0.98	-1.17	-0.59	-0.82	-1.01	-0.93	-0.95	-0.79
44	0.97	1.16	0.46	0.80	0.98	0.92	0.96	0.78
45	-0.64	-0.83	-0.52	-0.70	-0.70	-0.60	-0.68	-0.46
46	0.16	0.22	-0.22	-0.17	0.09	0.15	0.06	0.10
47	-1.39	-1.45	-0.65	-0.91	-1.35	-1.37	-1.26	-1.28
48	-0.03	-0.04	-0.31	-0.33	-0.11	-0.03	-0.13	-0.02
49	0.36	0.49	-0.10	0.02	0.30	0.33	0.27	0.23
50	-0.87	-1.07	-0.57	-0.79	-0.92	-0.82	-0.87	-0.68
$\Sigma(x)$	0.18	-2.70	-1.11	-1.97	-1.01	0.82	-0.59	2.46
$\Sigma(x^2)$	39.9808	43.9182	42.0127	51.0911	45.1523	39.0722	43.5785	35.7298

Note on Mr Quenouille's Edgeworth Type A transformation

By E. S. PEARSON

University College London

In representing a non-normal frequency function, $f(x)$, by a limited number of terms of an Edgeworth Type A expansion, statisticians have frequently used the form

$$f(x) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2} \left\{ 1 + \frac{\kappa_3}{6}(x^3 - 3x) + \frac{\kappa_4}{24}(x^4 - 6x^2 + 3) + \frac{\kappa_5^2}{72}(x^5 - 15x^3 + 45x - 15) \right\}. \quad (1)$$

This distribution, for example, formed the basis of Gayen's (1949, 1950) papers exploring the effect of departure from normality on several common statistical tests. It is also used by Srivastava in a paper printed elsewhere in this issue of *Biometrika*. It is important to realize, however, that the distributions of x_5 , x_6 and x_7 derived from a normally distributed x_1 through the Cornish-Fisher equations given by Quenouille on p. 179 above, will not correspond exactly to particular cases of the frequency function (1). It seemed of interest to examine the extent of this difference.

(i) For the case $\kappa_3 = 0.5$, $\kappa_4 = 0.0$, $f(x)$ of equation (1) becomes negative* at $x_5 = -2.53$, the negative area under the curve beyond this point amounting to -0.0028 . This means that the area above -2.53 is 1.0028 or is in excess by 3 in 1000. For the Quenouille transformation the critical point is at $x_5 = -1.86$. Thus there is inevitably some disagreement between the two distributions at the lower tail. It was found that at the upper tail there was good agreement† up to $x_5 = 3.2$, beyond which x_5 for distribution (1) begins to increase steadily over Quenouille's value given in his table on p. 182.

(ii) For the case $\kappa_3 = 0.0$, $\kappa_4 = 0.5$, $f(x)$ for (1) does not become negative and the two x_6 distributions scarcely differ until after $x_6 = \pm 3.4$.

(iii) For the case of $\kappa_3 = 0.5$, $\kappa_4 = 0.5$, the $f(x)$ of equation (1) becomes negative at $x_7 = -2.65$ and positive again at about $x_7 = -3.35$, but the effect of this anomaly is much smaller than for x_5 . Since these extreme negative and positive areas almost cancel out, the area under the curve above -2.65 is 0.99996 or only negligibly in error. In the lower tail, the values of x_7 obtained from (1) do not differ from those in Quenouille's table by more than 0.05 until $x_7 = -2.1$; beyond this point Quenouille's x_7 , for given x_1 , assumes increasingly greater negative values than the x_7 of (1). However, as the table below shows, this difference is unlikely to be noticeable in a sample of 1000.

In this table are given the expected frequencies for the two asymmetrical Edgeworth distributions of equation (1), expressed in terms of totals of 1000. These are compared with Quenouille's two samples tabled on p. 181. The only appreciable difference is at the lower end of the distribution of x_5 . For (1) there is an expected frequency of 8.6 in the interval -2.0 to -2.53 , while no observations are possible in this range using Quenouille's transformation. However, if the χ^2 test for goodness of fit is used to compare the sampled frequencies with the expectations tabled below, pooling tail groups as shown by the braces, it is found that

(a) for $\kappa_3 = 0.5$, $\kappa_4 = 0.0$: $\chi^2 = 6.29$, $\nu = 13$;

(b) for $\kappa_3 = 0.5$, $\kappa_4 = 0.5$: $\chi^2 = 15.65$, $\nu = 14$.

Neither of these values is significant. One must conclude therefore that except at the lower tail of the x_5 distribution it would be difficult to detect any difference between Quenouille's three Edgeworth distributions and those derived from the commonly used form of equation (1), even in samples as large as 1000.

REFERENCES

- BARTON, D. E. & DENNIS, K. E. R. (1952). *Biometrika*, **39**, 425-7.
 GAYEN, A. K. (1949). *Biometrika*, **36**, 353-69.
 GAYEN, A. K. (1950). *Biometrika*, **37**, 236-55, 399-408.

* This is to be expected from Barton & Dennis's (1952) investigation. $\beta_1 = \kappa_3^2$ and $\beta_2 = \kappa_4 + 3$ give a (β_1, β_2) point outside the region in which the $f(x)$ of (1) is positive definite.
 † That is, no difference between the corresponding values of over 0.05.

Comparison of Edgeworth Type A distributions

x	Case $\kappa_3 = 0.5, \kappa_4 = 0.0$, i.e. distribution of x_5		Case $\kappa_3 = 0.5, \kappa_4 = 0.5$, i.e. distribution of x_7	
	Expectation from (1)	Quenouille's sample	Expectation from (1)	Quenouille's sample
> 4.0	0.4	—	0.5	—
3.6 to 4.0	0.8	—	1.1	1
3.2 to 3.6	1.8	3	2.5	3
2.8 to 3.2	3.9	2	5.0	1
2.4 to 2.8	8.4	11	9.3	13
2.0 to 2.4	17.5	10	16.7	8
1.6 to 2.0	33.8	33	30.0	33
1.2 to 1.6	57.5	60	51.7	58
0.8 to 1.2	85.8	87	82.3	79
0.4 to 0.8	114.9	120	117.9	116
0.0 to 0.4	140.7	145	149.8	146
—0.4 to 0.0	155.6	154	164.7	167
—0.8 to —0.4	149.0	154	152.1	174
—1.2 to —0.8	117.4	117	113.9	112
—1.6 to —1.2	72.0	66	66.3	56
—2.0 to —1.6	31.9	38	28.1	26
—2.4 to —2.0	8.3	—	7.5	5
$< -2.4^*$	0.3	—	0.6	2
Total	1000.0	1000	1000.0	1000

* For equation (1), the lower limit for x_5 is -2.53 and for x_7 it is -2.65 .

MONOMIAL-MONOMIAL SYMMETRIC FUNCTION TABLES

BY JACK LEVINE

University of North Carolina, Raleigh

1. The present monomial-monomial (*MM*) Table is a continuation of the David-Kendall Tables (1949), (1951), (1953), (1955). All monomial products up to and including weight $w = 8$ are listed as linear functions of monomials. Thus, Table 8, section (iii) and (iv), gives

$$(21)^2(2) = (62) + 2(61^2) + 2(53) + 2(521) + 2(4^2) + 4(431) + 8(42^2) + 6(421^2) + 6(3^22) + 8(32^21) + 24(2^4) + 12(2^31^2).$$

2. The use of *MM* expansions in problems of distributions is illustrated by a theorem of MacMahon.

The number of distributions of n objects of specification $[m_1 m_2 \dots m_r]$ in r different boxes $B_1, \dots, B_r$, such that B_i contains p_i objects of specification $[m_{1i} m_{2i} \dots m_{ki}]$ is given by the coefficient $C_{m_1 \dots m_r}$ of the monomial symmetric function $(m_1 \dots m_l)$ in the expansion of the product

$$(m_{11} m_{21} \dots m_{k_1 1}) (m_{12} m_{22} \dots m_{k_2 2}) \dots (m_{1r} m_{2r} \dots m_{k_r r}) = \sum C_{q_1 \dots q_c} (q_1 \dots q_c),$$

the summation being over all partitions $[q_1 \dots q_c]$ of $n = \sum q_i = \sum p_i$.

3. Construction of the Tables was based on MacMahon's *D*-operator, the Tables of lower weight being used to build up those of higher weight. Checks were obtained by multiplication of factors in different orders.

The writer is indebted to Mr Charles N. Anderson for his valuable aid in the construction of these Tables.

Work on these Tables was supported by Office of Naval Research Contract Nonr 870(00).

REFERENCES

- DAVID, F. N. & KENDALL, M. G. (1949). *Biometrika*, **36**, 431.
 DAVID, F. N. & KENDALL, M. G. (1951). *Biometrika*, **38**, 435.
 DAVID, F. N. & KENDALL, M. G. (1953). *Biometrika*, **40**, 427.
 DAVID, F. N. & KENDALL, M. G. (1955). *Biometrika*, **42**, 223.

Table 2

$w = 2$	(2)	(1 ²)
(1) (1)	1	2

Table 3

$w = 3$	(3)	(21)	(1 ³)
(2) (1)	1	1	.
(1 ²) (1)	.	1	3
(1) ³	1	3	6

Table 4

$w = 4$	(4)	(31)	(2 ²)	(21 ²)	(1 ⁴)
(3) (1)	1	1	.	.	.
(21) (1)	.	1	2	2	.
(2) ²	1	.	2	.	.
(2) (1 ²)	.	1	.	1	.
(1 ³) (1)	.	.	.	1	4
(1 ²) ²	.	.	1	2	6
(2) (1) ²	1	2	2	2	.
(1 ²) (1) ²	.	1	2	5	12
(1) ⁴	1	4	6	12	24

Table 5

$w = 5$	(5)	(41)	(32)	(31 ²)	(2 ² 1)	(21 ²)	(1 ⁵)
(4) (1)	1	1	.	.	.	.	.
(31) (1)	.	1	1	2	.	.	.
(3) (2)	1	.	1	.	.	.	.
(3) (1 ²)	.	1	.	1	.	.	.
(2 ²) (1)	.	.	1	.	1	.	.
(21 ²) (1)	.	.	.	1	2	3	.
(21) (2)	.	1	1	.	2	.	.
(21) (1 ²)	.	.	1	2	2	3	.
(2) (1 ³)	.	.	.	1	.	1	.
(1 ⁴) (1)	.	.	.	.	.	1	5
(1 ³) (1 ²)	.	.	.	.	1	3	10
(3) (1) ²	1	2	1	2	.	.	.
(21) (1) ²	.	1	3	4	6	6	.
(2) ² (1)	1	1	2	.	2	.	.
(2) (1 ²) (1)	.	1	1	3	2	3	.
(1 ³) (1) ²	.	.	.	1	2	7	20
(1 ²) ² (1)	.	.	1	2	5	12	30
(2) (1) ³	1	3	4	6	6	6	.
(1 ²) (1) ³	.	1	3	7	12	27	60
(1) ⁵	1	5	10	20	30	60	120

Table 6

$w = 6$	(6)	(51)	(42)	(41 ²)	(3 ²)	(321)	(31 ³)	(2 ³)	(2 ² 1 ²)	(21 ⁴)	(1 ⁶)
(5) (1)	1	1	.	.	.	.	.	.	.	.	.
(41) (1)	.	1	1	2	.	.	.	.	.	.	.
(4) (2)	1	.	1	.	.	.	.	.	.	.	.
(4) (1 ²)	.	1	.	1	.	.	.	.	.	.	.
(32) (1)	.	.	1	.	2	1	.	.	.	.	.
(31 ²) (1)	.	.	.	1	.	1	.	.	.	.	.
(31) (2)	.	1	.	.	2	1	3	.	.	.	.
(31) (1 ²)	.	.	1	.	2	1	.	.	.	.	.
(3) ²	1	.	.	2	.	1	3	.	.	.	.
(3) (21)	.	1	1	.	2	.	.	.	.	.	.
(3) (1 ³)	.	.	.	1	.	1	.	.	.	.	.
(2 ² 1) (1)	.	.	.	.	.	1	.	3	.	.	.
(2 ²) (2)	.	.	1	.	.	.	.	3	2	.	.
(2 ²) (1 ²)	.	.	.	.	1	.	.	.	.	.	.
(21 ²) (1)	.	.	.	.	.	1	.	1	.	.	.
(21 ²) (2)	.	.	.	1	.	.	1	.	2	4	.
(21 ²) (1 ²)	.	.	.	.	.	1	.	2	2	.	.
(21) (1 ³)	.	.	1	2	2	2	3	6	4	6	.
(2) (1 ⁴)	.	.	.	.	.	1	3	.	2	4	.
(1 ⁵) (1)	.	.	.	.	.	.	1	.	.	1	.
(1 ⁴) (1 ²)	.	.	.	.	.	.	.	.	.	1	6
(1 ³) ²	.	.	.	.	.	.	.	.	1	4	15
(4) (1) ²	1	2	1	2	.	.	.	1	2	6	20
(31) (1) ²	.	1	2	4	2	3	6	.	.	.	.
(3) (2) (1)	1	1	1	.	2	.	.	.	.	.	.
(3) (1 ²) (1)	.	1	1	3	.	1	.	.	.	.	.
(2 ²) (1) ²	.	.	1	.	2	2	3	.	.	.	.
(21 ²) (1) ²	.	.	.	1	.	3	6	3	2	.	.
(21) (2) (1)	.	1	2	2	2	3	.	6	10	12	.
(21) (1 ²) (1)	.	.	1	2	2	5	9	6	10	12	.
(2) ² (1 ²)	1	.	3	.	.	.	.	6	.	.	.
(2) (1 ³) (1)	.	1	.	1	2	2	.	.	.	.	.
(2) (1 ²) ²	.	.	.	1	.	.	.	.	2	.	.
(1 ⁴) (1) ²	.	.	1	2	.	2	4	.	2	4	.
(1 ³) (1 ²) (1)	.	.	.	.	.	1	.	3	4	6	.
(1 ²) ³	.	.	.	.	.	.	1	.	2	9	30
(3) (1) ³	1	3	3	6	1	3	6	6	8	22	60
(21) (1) ³	.	1	4	6	2	3	6	6	15	36	90
(2) ² (1) ²	1	2	3	2	4	4	.	6	4	.	.
(2) (1 ²) (1) ²	.	1	2	5	2	6	.	.	.	.	.
(1 ³) ² (1) ²	.	.	.	1	.	3	12	6	10	12	.
(1 ²) ² (1) ³	.	.	1	2	.	8	10	6	18	48	120
(2) (1) ⁴	1	4	7	12	8	16	18	15	34	78	180
(1 ²) (1) ⁴	.	1	4	9	6	22	24	18	24	24	.
(1) ⁶	1	6	15	30	20	60	120	90	180	360	720

[illegible]

[illegible]

Table 8 (cont.)

$ic = 8$ (ii)	(41^4)	(3^22)	(3^21^2)	(32^21)	(321^2)	(31^5)	(2^4)	(2^21^2)	(2^21^4)	(21^6)	(1^8)
(7) (1)	.	.	.	.	.	.	.	.	.	.	.
(61) (1)	.	.	.	.	.	.	.	.	.	.	.
(6) (2)	.	.	.	.	.	.	.	.	.	.	.
(6) (1 ²)	.	.	.	.	.	.	.	.	.	.	.
(52) (1)	.	.	.	.	.	.	.	.	.	.	.
(51 ²) (1)	.	.	.	.	.	.	.	.	.	.	.
(51) (2)	.	.	.	.	.	.	.	.	.	.	.
(51) (1 ²)	.	.	.	.	.	.	.	.	.	.	.
(5) (3)	.	.	.	.	.	.	.	.	.	.	.
(5) (21)	.	.	.	.	.	.	.	.	.	.	.
(5) (1 ³)	.	.	.	.	.	.	.	.	.	.	.
(43) (1)	.	.	.	.	.	.	.	.	.	.	.
(421) (1)	.	.	.	.	.	.	.	.	.	.	.
(42) (2)	.	.	.	.	.	.	.	.	.	.	.
(42) (1 ²)	.	.	.	.	.	.	.	.	.	.	.
(41 ³) (1)	4	.	.	.	.	.	.	.	.	.	.
(41 ²) (2)	.	.	.	.	.	.	.	.	.	.	.
(41 ²) (1 ²)	6	.	.	.	.	.	.	.	.	.	.
(41) (3)	.	.	.	.	.	.	.	.	.	.	.
(41) (21)	.	.	.	.	.	.	.	.	.	.	.
(41) (1 ³)	4	.	.	.	.	.	.	.	.	.	.
(4) ²	.	.	.	.	.	.	.	.	.	.	.
(4) (31)	.	.	.	.	.	.	.	.	.	.	.
(4) (2 ²)	.	.	.	.	.	.	.	.	.	.	.
(4) (21 ²)	.	.	.	.	.	.	.	.	.	.	.
(4) (1 ⁴)	1	.	.	.	.	.	.	.	.	.	.
(3 ² 1) (1)	.	1	2	.	.	.	.	.	.	.	.
(3 ²) (2)	.	1	.	.	.	.	.	.	.	.	.
(3 ²) (1 ²)	.	.	1	.	.	.	.	.	.	.	.
(32 ²) (1)	.	2	.	1	.	.	.	.	.	.	.
(321 ²) (1)	.	.	2	2	3	.	.	.	.	.	.
(321) (2)	.	2	.	2	3	.	.	.	.	.	.
(321) (1 ²)	.	2	4	2	.	.	.	.	.	.	.
(32) (3)	.	2	.	.	.	.	.	.	.	.	.
(32) (21)	.	2	.	2	.	.	.	.	.	.	.
(32) (1 ³)	.	.	2	.	1	5	.	.	.	.	.
(31 ⁴) (1)	1	.	.	.	1	.	.	.	.	.	.
(31 ³) (2)	.	.	2	.	.	10	.	.	.	.	.
(31 ³) (1 ²)	4	.	.	1	3	.	.	.	.	.	.
(31 ²) (3)	.	.	2	.	.	.	.	.	.	.	.
(31 ²) (21)	.	.	.	2	3	.	.	.	.	.	.
(31 ²) (1 ³)	6	2	4	1	3	10	.	.	.	.	.
(31) (2 ³)	.	2	4	1	.	.	.	.	.	.	.
(31) (21 ²)	.	2	2	2	3	.	.	.	.	.	.
(31) (1 ⁴)	4	.	.	.	1	5	.	.	.	.	.
(3) (2 ² 1)	.	.	.	1	1	.	.	.	.	.	.
(3) (21 ³)	.	.	.	.	.	1	.	2	.	.	.
(3) (1 ⁵)	1	.	.	.	.	.	4	.	.	.	.
(2 ³ 1) (1)	.	.	.	1	.	.	4	1	.	.	.
(2 ³) (2)	.	.	.	.	.	.	.	3	4	.	.
(2 ³) (1 ²)	.	1	.	1	1	.	.	3	.	.	.
(2 ² 1 ³) (1)	.	.	.	1	.	.	6	6	6	.	.
(2 ² 1 ²) (2)	.	.	.	2	3	.	12	6	.	.	.
(2 ² 1 ²) (1 ²)	.	.	1	2	.	.	.	3	4	.	.
(2 ² 1) (21)	.	2	.	3	3	.	6	.	.	.	.
(2 ² 1) (1 ³)	.	1	2	.	.	.	.	3	1	.	.
(2 ²) (21 ²)	.	.	.	2	1	.	.	.	.	6	.
(2 ²) (1 ⁴)	.	1	.	.	.	.	.	.	2	.	.
(21 ⁵) (1)	.	.	.	.	1	1	.	3	8	15	.
(21 ⁴) (2)	.	.	.	.	.	5	.	6	.	.	.
(21 ⁴) (1 ²)	1	.	.	.	1	.	.	6	12	.	.
(21 ³) (21)	.	.	2	2	4	10	4	12	12	15	.
(21 ³) (1 ³)	4	.	.	1	3	.	12	3	8	6	.
(21 ³) ²	.	.	.	1	6	.	.	.	2	1	.
(21 ²) (1 ⁴)	6	2	4	4	3	10	.	.	.	1	8
(21) (1 ⁵)	.	.	.	.	1	1	.	.	.	.	.
(2) (1 ⁶)	.	.	.	.	.	.	.	.	.	6	28
(1 ⁷) (1)	.	.	.	.	.	.	.	1	4	15	56
(1 ⁶) (1 ²)	.	.	.	.	.	.	1	2	6	20	70
(1 ⁵) ²	.	.	.	.	.	.	.	.	.	.	.

Table 8 (cont.)

$w = 8$ (iii)	(8)	(71)	(62)	(61 ²)	(53)	(521)	(51 ³)	(4 ²)	(431)	(42 ²)	(421 ²)
(6) (1) ⁵	1	2	1	2	.	.	.	.	.	.	.
(51) (1) ⁴	.	1	2	4	1	3	6	.	.	.	.
(5) (2) (1)	1	1	1	.	1	1	.	.	.	.	.
(5) (1 ²) (1)	.	1	1	3	.	1	3	.	.	.	.
(42) (1) ²	.	.	1	.	2	2	.	2	2	2	2
(41 ²) (1) ²	.	.	.	1	.	2	6	.	1	2	5
(41) (2) (1)	.	1	1	2	1	1	.	2	2	2	2
(41) (1 ²) (1)	.	.	1	2	1	4	9	.	1	2	5
(4) (3) (1)	1	1	.	.	1	.	.	2	1	.	.
(4) (21) (1)	.	1	2	2	1	2	.	.	1	2	2
(4) (2) ²	1	.	2	.	.	.	.	2	.	2	.
(4) (2) (1 ²)	.	1	.	1	1	.	.	.	1	.	1
(4) (1 ³) (1)	.	.	.	1	1	1	4	.	.	.	1
(4) (1 ²) ²	.	.	1	2	.	2	6	.	.	1	2
(3 ²) (1) ²	.	.	.	.	1	.	.	2	2	.	.
(321) (1) ²	.	.	.	.	.	1	.	.	3	4	4
(32) (2) (1)	.	.	1	.	2	1	.	2	1	2	.
(32) (1 ²) (1)	.	.	.	.	1	1	.	2	4	2	3
(31 ²) (1) ²	.	.	.	.	.	.	1	.	.	.	2
(31 ²) (2) (1)	.	.	.	1	.	1	3	.	2	.	1
(31 ²) (1 ²) (1)	.	.	.	.	.	1	3	.	1	3	7
(31) (3) (1)	.	1	1	2	1	.	.	2	3	.	.
(31) (21) (1)	.	.	1	2	2	4	6	2	4	4	4
(31) (2) ²	.	1	.	.	3	2	.	.	1	.	.
(31) (2) (1 ²)	.	.	1	2	.	1	3	2	3	2	2
(31) (1 ³) (1)	.	.	.	.	.	1	3	.	1	2	6
(31) (1 ²) ²	.	.	.	.	1	3	6	.	2	4	10
(3) ² (2)	1	.	1	.	2	.	.	.	2	.	.
(3) ² (1 ²)	.	1	.	1	.	.	.	2	2	.	.
(3) (2 ²) (1)	.	.	1	.	1	1	.	.	.	1	.
(3) (21 ²) (1)	.	.	.	1	.	2	3	.	1	2	3
(3) (21) (2)	.	1	1	.	1	2	.	2	1	2	.
(3) (21) (1 ²)	.	.	1	2	2	4	6	.	1	2	3
(3) (2) (1 ²)	.	.	.	1	1	2	3	.	2	2	1
(3) (1 ⁴) (1)	.	.	.	.	.	.	1	.	1	.	1
(3) (1 ³) (1 ²)	.	.	.	.	.	.	1	.	.	.	3
(2 ³) (1) ²	.	.	.	.	.	1	3	.	.	1	.
(2 ² 1 ²) (1) ²	.	.	.	.	.	.	.	.	.	1	1
(2 ² 1) (2) (1)	.	.	.	.	.	.	.	.	.	.	2
(2 ² 1) (1 ³) (1)	.	.	.	.	.	1	.	.	1	3	2
(2 ²) (21) (1)	.	.	.	.	1	.	.	.	1	2	.
(2 ²) (2) ²	.	.	1	.	1	.	.	2	2	4	2
(2 ²) (2) (1 ²)	.	.	.	.	.	.	.	2	.	5	1
(2 ²) (1 ²) (1)	.	.	.	.	1	1	.	.	1	.	1
(2 ²) (1 ²) ²	.	.	.	.	.	.	.	.	1	.	1
(21 ⁴) (1) ²	.	.	.	.	.	.	.	1	2	2	2
(21 ³) (2) (1)	.	.	.	.	.	.	.	.	.	.	2
(21 ³) (1 ²) (1)	.	.	.	.	.	.	1	.	.	.	1
(21 ²) (21) (1)	.	.	.	.	.	.	.	.	.	.	8
(21 ²) (2) ²	.	.	.	1	.	1	3	.	3	4	3
(21 ²) (2) (1 ²)	.	.	.	.	.	1	.	.	2	.	.
(21) (2) (1 ³)	.	.	.	.	1	3	6	2	4	6	9
(21) (1 ⁴) (1)	.	.	.	.	.	1	3	.	1	.	2
(21) (1 ³) (1 ²)	.	.	.	.	.	.	.	.	.	.	1
(2) ² (1 ⁴)	.	.	.	.	.	.	.	.	.	.	5
(2) (1 ³) (1)	.	.	.	.	.	.	1	.	1	.	.
(2) (1 ⁴) (1 ²)	.	.	.	.	.	.	.	.	.	.	1
(2) (1 ³) ²	.	.	.	.	.	.	.	.	.	.	2
(1 ⁶) (1) ²	.	.	.	.	.	.	.	.	.	1	.
(1 ⁵) (1 ²) (1)	.	.	.	.	.	.	.	.	.	.	.
(1 ⁴) (1 ³) (1)	.	.	.	.	.	.	.	.	.	.	.
(1 ⁴) (1 ²) ²	.	.	.	.	.	.	.	.	.	.	.
(1 ³) ² (1 ²)	.	.	.	.	.	.	.	.	.	.	.

Table 8 (cont.)

$tc = 8$ (iv)	(41^4)	$(3^2 2)$	$(3^2 1^2)$	$(32^2 1)$	(321^2)	(31^5)	(2^4)	$(2^3 1^2)$	$(2^2 1^4)$	(21^6)	(1^8)
$(6) (1)^2$	.	.	.	.	.	.	.	.	.	.	.
$(51) (1)^2$	.	.	.	.	.	.	.	.	.	.	.
$(5) (2) (1)$	.	.	.	.	.	.	.	.	.	.	.
$(5) (1^2) (1)$	.	.	.	.	.	.	.	.	.	.	.
$(42) (1)^2$	.	.	.	.	.	.	.	.	.	.	.
$(41^2) (1)^2$	12	.	.	.	.	.	.	.	.	.	.
$(41) (2) (1)$	.	.	.	.	.	.	.	.	.	.	.
$(41) (1^2) (1)$	12	.	.	.	.	.	.	.	.	.	.
$(4) (3) (1)$	.	.	.	.	.	.	.	.	.	.	.
$(4) (21) (1)$	.	.	.	.	.	.	.	.	.	.	.
$(4) (2)^2$	.	.	.	.	.	.	.	.	.	.	.
$(4) (2) (1^2)$	.	.	.	.	.	.	.	.	.	.	.
$(4) (1^3) (1)$	4	.	.	.	.	.	.	.	.	.	.
$(4) (1^2)^2$	6	.	.	.	.	.	.	.	.	.	.
$(3^2) (1)^2$	.	1	2	.	6	.	.	.	.	.	.
$(321) (1)^2$	.	6	8	6	2	.	.	.	.	.	.
$(32) (2) (1)$	.	4	.	2	3	.	.	.	.	.	.
$(32) (1^2) (1)$	.	2	6	2	7	20	.	.	.	.	.
$(31^3) (1)^2$	8	.	2	2	3	.	.	.	.	.	.
$(31^2) (2) (1)$	.	2	6	2	3	.	.	.	.	.	.
$(31^2) (1^2) (1)$	18	2	4	5	12	30	.	.	.	.	.
$(31) (3) (1)$	.	2	4	6	6	.	.	.	.	.	.
$(31) (21) (1)$	.	6	8	6	.	.	.	.	.	.	.
$(31) (2)^2$	.	4	.	2	3	.	.	.	.	.	.
$(31) (2) (1^2)$	.	2	6	2	7	20	.	.	.	.	.
$(31) (1^3) (1)$	16	.	2	2	12	30	.	.	.	.	.
$(31) (1^2)^2$	24	2	4	5	.	.	.	.	.	.	.
$(3)^2 (2)$	.	2	.	.	.	.	.	.	.	.	.
$(3)^2 (1^2)$	.	.	2	.	.	.	.	.	.	.	.
$(3) (2^2) (1)$	.	2	.	1	.	.	.	.	.	.	.
$(3) (21^2) (1)$	.	.	2	2	3	.	.	.	.	.	.
$(3) (21) (2)$	.	2	.	2	3	.	.	.	.	.	.
$(3) (21) (1^2)$	.	2	4	2	1	.	.	.	.	.	.
$(3) (2) (1^3)$	.	.	2	.	1	5	.	.	.	.	.
$(3) (1^4) (1)$	5	.	.	.	3	10	.	2	.	.	.
$(3) (1^3) (1^2)$	10	.	.	1	3	.	4	15	12	.	.
$(2^3) (1)^2$	.	2	.	5	6	.	12	6	.	.	.
$(2^2 1^2) (1)^2$	.	.	2	4	.	.	12	15	12	.	.
$(2^2 1) (2) (1)$	.	2	.	9	9	.	12	6	.	.	.
$(2^2 1) (1^2) (1)$	.	5	6	.	.	.	12	.	.	.	.
$(2^2) (21) (1)$	.	4	.	5	.	.	.	3	4	.	.
$(2^2) (2)^2$	.	.	.	3	.	.	.	3	6	.	.
$(2^2) (2) (1^2)$	.	3	.	2	4	.	6	6	18	30	.
$(2^2) (1^3) (1)$	.	1	4	4	0	.	.	6	8	.	.
$(2^2) (1^2)^2$	.	2	6	.	3	10	.	6	60	.	.
$(21^4) (1)^2$	1	.	.	.	5	.	12	24	44	.	.
$(21^3) (2) (1)$	4	.	2	5	13	30	24	30	24	.	.
$(21^3) (1^2) (1)$	4	.	2	5	15	.	.	6	.	.	.
$(21^3) (21) (1)$	12	6	10	14	.	.	.	12	12	60	.
$(21^2) (2)^2$	.	2	.	4	.	.	12	24	44	90	.
$(21^2) (2) (1^2)$	6	2	6	6	0	40	12	45	72	.	.
$(21^2) (1^3) (1)$	6	2	4	8	30	60	24	12	24	.	.
$(21^2) (1^2)^2$	12	5	8	16	.	.	24	30	8	.	.
$(21^2) (2) (1)$	.	6	.	8	18	.	.	6	18	30	.
$(21^2) (1^2) (1)$	12	10	14	18	7	25	12	24	44	60	.
$(21) (2) (1^3)$	4	2	6	4	9	50	.	.	2	6	.
$(21) (1^4) (1)$	4	.	2	9	22	.	.	.	.	.	.
$(21) (1^3) (1^2)$	12	2	6	.	2	6	.	.	.	.	.
$(2)^2 (1^4) (1)$	1	.	2	.	1	.	.	3	8	15	.
$(2) (1^5) (1)$	1	.	.	.	.	.	.	6	12	20	.
$(2) (1^4) (1^2)$	4	.	.	1	4	15	4	.	2	13	56
$(1^6) (1)^2$	0	.	.	2	6	20	.	.	3	51	168
$(1^5) (1^2)$	.	.	.	.	1	5	.	11	32	95	280
$(1^4) (1^3) (1)$	.	.	.	1	3	10	6	18	53	150	420
$(1^4) (1^2)^2$	.	.	1	2	7	30	12	31	80	210	500
$(1^3)^2 (1^2)$	.	1	2	5	12	.	.	.	.	.	.

Table 8 (cont.)

$w = 8$ (v)	(8)	(71)	(62)	(61 ²)	(53)	(521)	(51 ³)	(4 ²)	(431)	(42 ²)	(421 ²)
(5) (1) ³	1	3	3	6	1	3	6	.	.	.	.
(41) (1) ³	.	1	3	6	3	9	18	2	4	6	12
(4) (2) (1) ²	1	2	2	2	2	2	.	2	2	2	2
(4) (1 ²) (1) ²	.	1	2	5	1	5	12	.	1	2	5
(32) (1) ³	.	.	1	.	4	3	.	6	9	6	6
(31 ²) (1) ³	.	.	.	1	.	3	9	.	4	6	15
(31) (2) (1) ²	.	1	2	4	3	4	6	4	7	4	4
(31) (1 ²) (1) ²	.	.	1	2	2	7	15	2	7	10	22
(3) ² (1) ²	1	2	1	2	2	.	.	4	4	.	.
(3) (21) (1) ²	.	1	3	4	3	6	6	2	5	6	6
(3) (2) ² (1)	1	1	2	.	3	2	.	2	1	2	.
(3) (2) (1 ²) (1)	.	1	1	3	1	2	3	2	4	2	3
(3) (1 ³) (1) ²	.	.	.	1	.	2	7	.	1	2	7
(3) (1 ²) ² (1)	.	.	1	2	1	5	12	.	2	5	12
(2 ² 1) (1) ³	.	.	.	.	.	1	.	.	3	7	6
(2 ²) (2) (1) ²	.	.	1	.	2	2	.	2	2	5	2
(2 ²) (1 ²) (1) ²	.	.	.	.	1	1	.	2	5	4	5
(21 ²) (1) ³	.	.	.	.	.	1	1	.	.	4	4
(21 ²) (2) (1) ²	.	.	.	1	.	3	6	.	4	6	11
(21 ²) (1 ²) (1) ²	.	.	.	.	.	1	3	.	3	7	14
(21) ² (1) ²	.	.	1	2	4	8	12	6	14	20	24
(21) (2) ² (1)	.	1	2	2	3	4	.	4	5	10	6
(21) (2) (1 ²) (1)	.	.	1	2	2	5	9	2	7	8	12
(21) (1 ³) (1) ²	.	.	.	.	.	1	3	2	3	4	12
(21) (1 ²) ² (1)	.	.	.	.	1	3	6	2	8	12	24
(2) ⁴	1	.	4	.	.	.	.	6	.	12	.
(2) ³ (1 ²)	.	1	.	1	3	3	.	.	3	.	3
(2) ² (1 ²) (1)	.	.	.	1	3	1	4	.	2	.	3
(2) ² (1 ²) ²	.	.	1	2	.	2	6	2	4	5	2
(2) (1 ⁴) (1) ²	.	.	.	.	.	1	.	.	.	.	.
(2) (1 ³) (1 ²) (1)	.	.	.	.	.	1	3	.	1	3	8
(2) (1 ²) ²	.	.	.	.	1	3	6	.	3	6	15
(1 ³) (1) ²	.	.	.	.	.	.	.	.	.	.	1
(1 ⁴) (1 ²) (1) ²	.	.	.	.	.	.	.	.	.	1	2
(1 ³) ² (1) ²	.	.	.	.	.	.	.	.	1	2	5
(1 ²) ⁴	.	.	.	.	.	.	.	.	.	.	12
(4) (1) ⁴	1	4	6	12	4	12	24	1	4	6	12
(31) (1) ⁴	.	1	4	8	7	18	36	2	4	6	12
(3) (2) (1) ³	1	3	4	6	5	6	6	8	21	24	48
(3) (1 ³) (1) ³	.	1	3	7	3	12	27	6	8	12	27
(2 ²) (1) ⁴	.	.	1	.	4	4	.	6	12	13	12
(21 ²) (1) ⁴	.	.	.	1	.	5	12	.	10	20	39
(21) (2) (1) ³	.	1	4	6	7	14	18	8	19	26	30
(21) (1 ²) (1) ³	.	.	1	2	4	11	21	6	23	32	60
(2) ² (1) ²	1	2	4	2	6	6	.	6	6	12	6
(2) ² (1 ²) (1) ²	.	1	2	5	3	7	12	4	11	10	15
(2) (1 ³) (1) ³	.	.	.	1	.	3	10	.	4	6	19
(2) (1 ²) ² (1) ²	.	.	1	2	2	8	18	2	10	17	36
(1 ⁴) (1) ⁴	.	.	.	.	.	.	1	.	.	.	4
(1 ³) (1 ²) (1) ³	.	.	.	.	.	1	3	.	3	7	18
(1 ²) ³ (1) ²	.	.	.	.	1	3	6	.	11	18	39
(3) (1) ⁵	1	5	10	20	11	30	60	2	25	30	60
(21) (1) ⁵	.	1	6	10	15	36	60	10	65	90	150
(2) ² (1) ⁴	1	4	8	12	12	20	24	20	28	32	36
(2) (1 ³) (1) ⁴	.	1	4	9	7	23	48	14	31	44	87
(1 ³) (1) ⁵	.	.	.	1	.	5	16	8	10	20	55
(1 ²) ² (1) ⁴	.	.	1	2	4	14	30	.	10	20	114
(2) (1) ⁶	1	6	16	30	26	66	120	6	32	53	210
(1 ²) (1) ⁵	.	1	6	13	15	51	108	30	90	120	315
(1) ⁸	1	8	28	56	56	168	336	20	95	150	840

Table 8 (cont.)

$w = 8$ (vi)	(41^4)	$(3^2 2)$	$(3^2 1^2)$	$(32^2 1)$	(321^2)	(31^5)	(2^4)	$(2^3 1^2)$	$(2^2 1^4)$	(21^6)	(1^8)
$(5)(1)^3$	.	.	.	.	.	.	.	.	.	.	.
$(41)(1)^3$	24	.	.	.	.	.	.	.	.	.	.
$(4)(2)(1)^2$	.	.	.	.	.	.	.	.	.	.	.
$(4)(1^2)(1)^2$	12	.	.	.	.	.	.	.	.	.	.
$(32)(1)^3$	.	8	12	6	6	.	.	.	.	.	.
$(31^2)(1)^3$	36	6	14	12	27	60	.	.	.	.	.
$(31)(2)(1)^2$	.	8	12	6	6	.	.	.	.	.	.
$(31)(1^2)(1)^2$	48	6	14	12	27	60	.	.	.	.	.
$(3)^2(1)^2$	.	2	4	.	.	.	.	.	.	.	.
$(3)(21)(1)^2$	.	6	8	6	6	.	.	.	.	.	.
$(3)(2)^2(1)$	.	4	.	2	.	.	.	.	.	.	.
$(3)(2)(1^2)(1)$	.	2	6	2	3	.	.	.	.	.	.
$(3)(1^3)(1)^2$	20	.	2	2	7	20	.	.	.	.	.
$(3)(1^2)^2(1)$	30	2	4	5	12	30	.	.	.	.	.
$(2^2 1)(1)^2$	.	12	12	22	18	.	36	36	24	.	.
$(2^2)(2)(1)^2$	.	6	.	6	.	.	12	6	.	.	.
$(2^2)(1^2)(1)^2$	.	7	12	11	12	.	12	15	12	.	.
$(21^3)(1)^2$	12	.	6	12	31	60	24	54	96	120	.
$(21^2)(2)(1)^2$	12	6	12	16	18	.	24	30	24	.	.
$(21^2)(1^2)(1)^2$	30	12	22	38	69	120	60	102	156	180	.
$(21)^2(1)^2$	.	.	.	.	.	.	72	72	48	.	.
$(21)(2)^2(1)$	24	26	28	44	36	.	24	12	.	.	.
$(21)(2)(1^2)(1)$	.	8	.	10	.	.	24	30	24	.	.
$(21)(1^3)(1)^2$	12	12	18	20	21	.	24	54	96	120	.
$(21)(1^2)^2(1)$	28	6	18	22	51	100	24	54	96	180	.
$(21)(1^2)(1^2)(1)$	48	16	32	45	84	150	24	60	102	156	.
$(2)^3$	.	.	.	.	.	.	.	6	.	.	.
$(2)^2(1^2)$	.	6	.	6	.	.	.	6	8	.	.
$(2)^2(1^2)(1)$	4	2	8	4	8	.	12	12	12	.	.
$(2)^2(1^2)^2$	6	4	12	8	12	.	.	6	18	30	.
$(2)(1^4)(1)^2$	9	.	2	2	10	30	.	.	.	.	.
$(2)(1^3)(1^2)(1)$	22	2	6	10	25	60	12	24	44	60	.
$(2)(1^2)^2$	36	7	12	21	42	90	24	45	72	90	336
$(1^4)(1)^3$	1	.	.	.	3	16	.	6	30	108	840
$(1^4)(1^2)(1)^2$	4	.	2	5	18	55	12	39	114	315	1120
$(1^3)^2(1)^2$	6	2	4	12	30	80	28	68	172	440	1680
$(1^3)(1^2)^2(1)$	12	5	12	24	58	140	48	117	284	690	2520
$(1^3)^4$	24	12	28	48	108	240	90	204	468	1080	.
$(4)(1)^4$	24	.	.	.	.	.	.	.	.	.	.
$(31)(1)^4$	96	20	40	30	60	120	.	.	.	.	.
$(3)(2)(1)^3$	.	8	12	6	6	.	.	.	.	.	.
$(3)(1^2)(1)^3$	60	6	14	12	27	60	.	36	36	24	.
$(2^2)(1)^4$	.	20	24	28	24	.	144	234	336	360	.
$(21^2)(1)^4$	72	30	56	92	156	240	72	72	48	360	.
$(21)(2)(1)^3$	24	32	36	50	42	.	144	234	336	.	.
$(21)(1^2)(1)^3$	108	44	82	110	189	300	24	12	.	.	.
$(2)^2(1^2)$	.	12	.	12	.	.	24	30	24	120	.
$(2)^2(1^2)(1)^2$	12	14	24	22	24	.	24	54	96	180	.
$(2)(1^3)(1)^3$	48	6	20	24	58	120	24	102	156	660	1680
$(2)(1^2)^2(1)^2$	78	18	36	50	96	180	24	84	246	.	.
$(1^4)(1)^4$	17	.	6	12	46	140	.	.	.	.	.
$(1^3)(1^2)(1)^3$	46	12	30	58	141	340	108	258	612	1440	3360
$(1^2)^3(1)^2$	84	31	68	117	258	570	204	453	1008	2250	5040
$(3)(1)^5$	120	20	40	30	60	120	.	.	.	720	.
$(21)(1)^5$	240	120	200	270	420	600	360	540	720	48	.
$(2)^2(1)^4$	24	40	48	56	48	.	72	234	336	360	6720
$(2)(1^2)(1)^4$	168	50	96	122	216	360	144	570	1320	3000	10080
$(1^3)(1^3)(1)^4$	140	30	80	140	340	800	240	570	2172	4680	.
$(1^2)^2(1)^4$	246	80	172	284	612	1320	468	1008	720	720	20160
$(2)(1)^6$	360	140	240	300	480	720	360	540	4680	.	.
$(1^2)(1)^6$	660	210	440	690	1440	3000	1080	2250	.	.	.
$(1)^8$	1680	560	1120	1680	3360	6720	2520	5040	10080	20160	40320

MISCELLANEA

Approximate linearization of the incomplete β -function

BY A. W. KIMBALL AND E. LEACH

Oak Ridge National Laboratory, Tennessee†

1. INTRODUCTION

The incomplete β -function has been a favourite subject with statisticians for many years because it arises frequently in our work and because it presents problems in evaluation. Several tables of this function and related functions have been constructed and many approximations have been derived. Wise (1950) gives an excellent account of various approximations in terms of percentage points of the χ^2 -distribution and provides methods for computing percentage errors. From his paper, and other literature which is too voluminous to review here, one can ascertain that most efforts have been directed toward the numerical evaluation of either the function itself or of one of its arguments, usually in connexion with the determination of Type I or Type II errors for test statistics, the construction of confidence intervals, or the determination of sample sizes. When one attempts to use existing formulae and tables, one is immediately impressed by the fact that no single table or formula is sufficient for every purpose. Tables may be unsatisfactory either because the range of the arguments is too narrow or because interpolation is required. The approximation formulae may be unsatisfactory because they are derived for significance tests or confidence intervals and hence are restricted to percentage points usually near the tails of the distribution.

The approximation reported in this paper resulted from the need for a single representation of the incomplete β -function that would be valid over wide ranges of the function and the three arguments. In radiation mortality studies, an approach known popularly as 'target theory' predicts that, for some organisms, the probability (M) of death from a dose (D) of radiation is given by

$$M = \sum_{i=p}^{p+q-1} C_i^{p+q-1} x^i (1-x)^{p+q-1-i} = I_x(p, q),$$

where p and q are unknown parameters, x is a function of D , and $I_x(p, q)$ is Pearson's (1934) notation for the incomplete beta function, i.e.

$$I_x(p, q) = \int_0^x t^{p-1} (1-t)^{q-1} dt \bigg/ \int_0^1 t^{p-1} (1-t)^{q-1} dt.$$

From observations on M and D , regression estimates may be obtained, although the use of $I_x(p, q)$ as an expected value presents formidable difficulties. To avoid these difficulties, various attempts were made to find a transformation that would provide an approximate linear relationship between functions of the dependent and independent variables. To be of any value the approximation had to be satisfactory over almost the entire range of $I_x(p, q)$ and had to be valid for values of p, q which were unknown, and for values of x which varied over an unknown range. A search of the literature revealed that, for reasons already cited, no available single formula would suffice. However, by combining two formulae found in the literature and by making one modification, we arrived at a linear transformation that proved adequate for the regression problem. Although it is less accurate than some approximations for the purpose of evaluating $I_x(p, q)$, it has obvious advantages for use as an expected value in the curve-fitting problem.

2. THE APPROXIMATION

In connexion with a sample size problem, Scheffé & Tukey (1944) give, as an approximate solution to the equation $I_x(p, q) = \alpha$,

$$(p+q-1) = \frac{1}{4} \chi_\alpha^2 \left(\frac{1+x}{1-x} \right) + \frac{1}{2}(q-1), \quad (1)$$

where χ_α^2 is the 100 α % point of the χ^2 -distribution with $2q$ degrees of freedom. They indicate that for $0.005 \leq \alpha \leq 0.10$ and $x \geq 0.9$, the error in this solution is less than 0.1 %. It also does well for larger

† Operated by Union Carbide Corporation for the U.S. Atomic Energy Commission.

values of α but becomes inaccurate as x is decreased. We have found that the approximation can be extended to smaller values of x if the first term is multiplied by

$$\left(\frac{21}{23} + \frac{1}{19x} + \frac{1}{3}e^{-5/(2x)}\right)^{-1}. \quad (2)$$

The following modification of the Scheffé-Tukey approximation has been obtained by Tukey (1958)

$$\chi_{\alpha}^2 = 2q(1 - \sqrt{x}) + (2p - 1)(1/\sqrt{x} - \sqrt{x}). \quad (3)$$

Computations over the range of arguments in Table 1 show that this new form is remarkably superior to the original approximation. Unfortunately, it does not permit linearization of the regression problem.

Wilson & Hilferty (1931) give the following approximation to χ_{α}^2 ,

$$\chi_{\alpha}^2 = 2q \left(1 - \frac{1}{9q} + y_{1-\alpha} \sqrt{\frac{1}{9q}}\right)^3, \quad (4)$$

where $y_{1-\alpha}$ is defined by

$$\int_{-\infty}^{y_{1-\alpha}} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}t^2} dt = \alpha.$$

According to the results of Goldberg & Levine (1946), the Wilson-Hilferty approximation, when compared with the three others, does best over a wider range of arguments and hence was chosen for use in our regression problem.

By combining (4) with (1) as modified by (2), one obtains

$$y_{\alpha} = A + BX, \quad (5)$$

where

$$A = \frac{9q-1}{3\sqrt{q}}, \quad B = -3\sqrt{q} \left[1 + \frac{2p-1}{q}\right]^{\frac{1}{2}},$$

$$X = \left[\left(\frac{1-x}{1+x}\right) \left(\frac{21}{23} + \frac{1}{19x} + \frac{1}{3}e^{-5/(2x)}\right) \right]^{\frac{1}{2}}.$$

Thus, for any x , p and q , y_{α} can be computed readily and the approximation,

$$\hat{f} = I_x(p, q) = \int_{-\infty}^{y_{\alpha}} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}t^2} dt, \quad (6)$$

obtained from tables of the normal distribution. The advantage of this approximation in the aforementioned regression problem should be obvious. Not only is the problem linear, but the dependent variable plus five is a probit, and the literature abounds with computing aids for probit analysis.

By combining (3) with (4), one obtains

$$y_{\alpha} = A - 3\sqrt{q} \left[(1-x) + \frac{2p-1}{2q} \left(\frac{1-x}{\sqrt{x}}\right) \right]^{\frac{1}{2}}. \quad (7)$$

In Table 1, the approximation obtained from (7) is identified as I^* . For evaluating $I_x(p, q)$, this form is

definitely superior to (5), but it is of little use in the regression problem.

In Table 1, three approximations are compared with exact values as obtained from Pearson's (1934) tables and from the Harvard tables (1955). The range covered is $2 \leq p \leq 500$ and $1 \leq q \leq 451$ for each of

four values of x chosen so that no interpolation would be required to obtain the exact values and so that $I_x(p, q)$ would have values near 0.05, 0.30, 0.70 and 0.95. To conserve space, the values of x are not shown.

The ranges of p and q correspond to sample sizes between 2 and 951 for the related binomial distribution.

Since, for large values of p and q , the direct normal approximation to $I_x(p, q)$ would be expected to do quite well, Table 1 also contains the approximation obtained by taking

$$y_{\alpha} = \frac{p(1-x) - x(q-1) - 0.5}{[(p+q-1)x(1-x)]^{\frac{1}{2}}}. \quad (8)$$

In Table 1, the approximation obtained from (8) is identified as I' .

3. GOODNESS OF THE APPROXIMATIONS

Inspection of Table 1 shows that $\hat{f}$ does generally better than I' for values of $p \leq 80$ and $q \leq 41$ except where p and q are about equal. This latter case corresponds to a probability near one-half for the binomial distribution in which case the normal approximation provides good accuracy even for relatively small

Table 1†. *Approximations to $I_x(p, q)$* I = exact value I' = approximation from (6) and (8) $\hat{I}$ = approximation from (5) and (6) I^* = approximation from (6) and (7)(Numbers in parentheses are errors to the nearest *third* decimal.)

p	q		x_1	x_2	x_3	x_4
2	1	I	0.04840	0.30250	0.70560	0.94090
		$\hat{I}$	·05109 (+ 3)	·30871 (+ 6)	·71294 (+ 7)	·93178 (− 9)
		I'	·03519 (− 13)	·28483 (− 18)	·63577 (− 70)	·96591 (+ 25)
		I^*	·04724 (− 1)	·31094 (+ 8)	·71291 (+ 7)	·93149 (− 9)
2	2	I	0.05331	0.29549	0.70451	0.94669
		$\hat{I}$	·04478 (− 8)	·28806 (− 7)	·70235 (− 2)	·94316 (− 4)
		I'	·03617 (− 17)	·30672 (+ 11)	·69328 (− 11)	·96383 (+ 17)
		I^*	·05120 (− 2)	·30762 (+ 12)	·71137 (+ 7)	·94360 (− 3)
5	2	I	0.05097	0.30065	0.70441	0.95408
		$\hat{I}$	·05247 (+ 2)	·30372 (+ 3)	·70941 (+ 5)	·95176 (− 2)
		I'	·05074 (0)	·28728 (− 13)	·67231 (− 32)	·97499 (+ 21)
		I^*	·04987 (− 1)	·30253 (+ 2)	·70829 (+ 4)	·95134 (− 3)
5	4	I	0.05054	0.29985	0.70640	0.95244
		$\hat{I}$	·04715 (− 3)	·29179 (− 8)	·70221 (− 4)	·95055 (− 2)
		I'	·04470 (− 6)	·30007 (0)	·69808 (− 8)	·96282 (+ 10)
		I^*	·04887 (− 2)	·30346 (+ 4)	·70939 (+ 3)	·95113 (− 1)
12	2	I	0.04731	0.29205	0.72063	0.94363
		$\hat{I}$	·04884 (+ 2)	·29602 (+ 4)	·72649 (+ 6)	·94207 (− 2)
		I'	·05688 (+ 10)	·27212 (− 20)	·68092 (− 40)	·96444 (+ 21)
		I^*	·04661 (0)	·29187 (0)	·72385 (+ 3)	·94142 (− 2)
12	6	I	0.05167	0.29231	0.70109	0.94521
		$\hat{I}$	·05197 (0)	·29274 (0)	·70258 (+ 2)	·94505 (0)
		I'	·05246 (+ 1)	·28709 (− 5)	·69045 (− 11)	·95403 (+ 9)
		I^*	·05093 (− 1)	·29312 (+ 1)	·70265 (+ 2)	·94469 (0)
12	10	I	0.04811	0.29888	0.69144	0.94519
		$\hat{I}$	·04280 (− 5)	·28367 (− 15)	·67950 (− 12)	·94193 (− 3)
		I'	·04591 (− 2)	·29868 (0)	·68792 (− 4)	·94949 (+ 4)
		I^*	·04701 (− 1)	·30057 (+ 2)	·69304 (+ 2)	·94450 (− 1)
30	5	I	0.04909	0.29168	0.68274	0.94936
		$\hat{I}$	·05149 (+ 2)	·29782 (+ 6)	·68925 (+ 6)	·95013 (+ 1)
		I'	·05657 (+ 8)	·27948 (− 12)	·66150 (− 21)	·96217 (+ 13)
		I^*	·04878 (0)	·29138 (0)	·68408 (+ 1)	·94894 (0)
30	15	I	0.05268	0.29271	0.72342	0.95542
		$\hat{I}$	·05389 (+ 1)	·29295 (0)	·72409 (+ 1)	·95540 (0)
		I'	·05413 (+ 1)	·28917 (− 4)	·71840 (− 5)	·96030 (+ 5)
		I^*	·05227 (0)	·29312 (0)	·72409 (+ 1)	·95518 (0)
30	25	I	0.04282	0.29726	0.69384	0.94300
		$\hat{I}$	·03596 (− 7)	·27380 (− 24)	·67286 (− 21)	·93690 (− 6)
		I'	·04216 (− 1)	·29683 (0)	·69210 (− 2)	·94495 (+ 2)
		I^*	·04215 (− 1)	·29810 (+ 1)	·69469 (+ 1)	·94255 (0)

Table 1 (cont.)

p	q		x_1	x_2	x_3	x_4
80	11	I	0.05292	0.26911	0.71251	0.95032
		$\hat{I}$	.05644 (+ 4)	.27806 (+ 9)	.72087 (+ 8)	.95220 (+ 2)
		I'	.05892 (+ 6)	.26175 (- 7)	.70092 (- 12)	.95864 (+ 8)
		I^*	.05277 (0)	.26882 (0)	.71307 (+ 1)	.95021 (0)
80	41	I	0.05198	0.30621	0.66294	0.94966
		$\hat{I}$	.05184 (0)	.30574 (0)	.66270 (0)	.94957 (0)
		I'	.05318 (+ 1)	.30380 (- 2)	.69545 (+ 32)	.95210 (+ 2)
		I^*	.05176 (0)	.30644 (0)	.66336 (0)	.94949 (0)
80	71	I	0.04292	0.31229	0.68943	0.95914
		$\hat{I}$	.03102 (- 12)	.26657 (- 46)	.64480 (- 45)	.94845 (- 11)
		I'	.04270 (0)	.31207 (0)	.68877 (- 1)	.95994 (+ 1)
		I^*	.04250 (0)	.31286 (+ 1)	.69000 (+ 1)	.95881 (0)
200	41	I	0.05543	0.33028	0.65056	0.95911
		$\hat{I}$	.06088 (+ 6)	.34758 (+ 17)	.66684 (+ 16)	.96240 (+ 3)
		I'	.05833 (+ 3)	.32504 (- 5)	.64422 (- 6)	.96287 (+ 4)
		I^*	.05536 (0)	.33029 (0)	.65079 (0)	.95909 (0)
200	121	I	0.05711	0.31223	0.73254	0.96066
		$\hat{I}$	.05237 (- 5)	.29822 (- 14)	.72022 (- 12)	.95763 (- 3)
		I'	.05770 (+ 1)	.31107 (- 1)	.73147 (- 1)	.96175 (+ 1)
		I^*	.05694 (0)	.31254 (0)	.73274 (0)	.96051 (0)
200	181	I	0.03961	0.27940	0.72157	0.96143
		$\hat{I}$	.02318 (- 16)	.20972 (- 70)	.64741 (- 74)	.94193 (- 20)
		I'	.03956 (0)	.27930 (0)	.72129 (0)	.96179 (0)
		I^*	.03935 (0)	.27969 (0)	.72184 (0)	.96120 (0)
500	101	I	0.02152	0.21393	0.69515	0.97166
		$\hat{I}$	.02607 (+ 5)	.23660 (+ 23)	.71934 (+ 24)	.97562 (+ 4)
		I'	.02328 (+ 2)	.21273 (- 1)	.69186 (- 3)	.97389 (+ 2)
		I^*	.02150 (0)	.21391 (0)	.69533 (0)	.97162 (0)
500	301	I	0.02364	0.20249	0.63018	0.93511
		$\hat{I}$	.02029 (- 3)	.18566 (- 17)	.60799 (- 22)	.92779 (- 7)
		I'	.02403 (0)	.20225 (0)	.62912 (- 1)	.93569 (+ 1)
		I^*	.02357 (0)	.20236 (0)	.63030 (0)	.93507 (0)
500	451	I	0.05592	0.36063	0.60280	0.93303
		$\hat{I}$	.02576 (- 30)	.24218 (- 118)	.47013 (- 133)	.88040 (- 53)
		I'	.05594 (0)	.36048 (0)	.60258 (0)	.93319 (0)
		I^*	.05578 (0)	.36096 (0)	.60313 (0)	.93294 (0)

† Abridged from a larger table available from the author on request.

sample sizes. On the other hand, the approximation given by $\hat{I}$ is not too bad for these cases, particularly for purposes of regression, and its accuracy tends to improve as x increases. As expected, for $p > 80$ and $q > 41$, the normal approximation is uniformly better than $\hat{I}$ and in fact is good enough for almost any practical purpose including the regression problem.

The approximation I^* is surprisingly accurate over the ranges examined and even competes successfully with I' for large p and q . Although $\hat{I}$ does not have the uniform accuracy of I^* , or of I' for large p and q , it is definitely preferable for use in the regression problem in the lower ranges of p and q . Perhaps it should be emphasized that all of these approximations may be used, without the need for interpolation, for integral or fractional values of p and q , the only tables required being the readily available tables of the normal distribution.

REFERENCES

- GOLDBERG, HENRY & LEVINE, HARRIET (1946). Approximate formulas for the percentage points and normalization of t and χ^2 . *Ann. Math. Statist.* **17**, 216–25.
- HARVARD COMPUTATION LABORATORY (1955). *Tables of the Cumulative Binomial Probability Distribution*. Cambridge, Massachusetts: Harvard University Press.
- PEARSON, KARL (1934). *Tables of the Incomplete Beta Function*. Cambridge, England: The University Press.
- SCHEFFÉ, H. & TUKEY, J. W. (1944). A formula for sample sizes for population tolerance limits. *Ann. Math. Statist.* **15**, 217.
- TUKEY, J. W. (1958). Private communication.
- WILSON, EDWIN B. & HILFERTY, MARGARET M. (1931). The distribution of chi-square. *Proc. Nat. Acad. Sci.* **17**, 684–8.
- WISE, M. E. (1950). The incomplete beta function as a contour integral and a quickly converging series for its inverse. *Biometrika*, **37**, 208–18.

Sequential occupancy

By D. E. BARTON AND F. N. DAVID

University College London

INTRODUCTION

In previous papers (Barton & David, 1959*a, b*) we have discussed the distribution of the number of empty compartments when n balls are randomly and independently thrown into N identical compartments. Various restrictions on the compartments have been envisaged. The alternative problem where the number of empty compartments is fixed and the number of balls becomes the random variable is not without interest. The classical case corresponding to De Moivre's differences of zero distribution is, of course, well known, but we think that this is not true of the other cases discussed below.

The actual distributions are simply and easily obtained. Suppose k zero compartments and let the distribution of n , the number of balls required to fill just $N-k$ compartments, be defined. Then the probability of n is equal to the probability that $n-1$ balls will fill $N-k-1$ compartments times the probability that one further ball falls in an empty compartment. This argument will hold whether there is restriction on the compartment size or not, and the sequential distributions thus follow from our previous work (1959*a*). The generalization of Pólya's scheme has also been noted previously.

The moments of the distribution of n are not easy to find unless the actual probability process is investigated (see for example Feller, 1957, p. 210). We discuss here methods of obtaining these moments.

CLASSICAL CASE WITH NO RESTRAINT ON BALLS OR COMPARTMENTS

The probability distribution of n is

$$p(n) = \frac{1}{N^{n-1}} {}^N C_{k+1} \Delta^{N-k-1} 0^{n-1} \frac{k+1}{N} = \frac{1}{N^{n-1}} {}^{N-1} C_k \Delta^{N-k-1} 0^{n-1} \quad (n = N-k, N-k+1, \dots, \infty),$$

a fact which is well known. The moments of n are straightforward. Suppose the dropping of the balls is going on and we break in at the point when the $(N-t)$ th box, empty before, receives one ball leaving $t (\leq k)$ boxes empty. Let there be n_t balls in the $(N-t)$ boxes. Let the further number of balls dropped

up to and including the occasion when one further empty box receives one ball be α_{N-t} . Then the total number of balls necessary to achieve $N-k$ filled and k empty is

$$n = \sum_{t=k+1}^N \alpha_{N-t},$$

with $\alpha_0 = 1$. α_{N-t} is a sequential binomial variable with the trials stopping when just one success is achieved. Accordingly

$$K_{[r]}(\alpha_{N-t}) = (r-1)! \left(\left(\frac{q}{p} \right)^r + (-1)^{r-1} \right)$$

with

$$p = \frac{t}{N}, \quad q = \frac{N-t}{N},$$

or

$$K_{[r]}(\alpha_{N-t}) = (r-1)! \left(\left(\frac{N-t}{t} \right)^r + (-1)^{r-1} \right).$$

In this classical scheme the α 's are mutually independent and consequently

$$K_{[r]}(n) = (r-1)! \left((N-k)(-1)^{r-1} + \sum_{t=k+1}^N \left(\frac{N-t}{t} \right)^r \right),$$

or

$$K_{[r]}(n) = (r-1)! \left((N-k)(-1)^{r-1} + \sum_{j=0}^{N-k-1} \left(\frac{j}{N-j} \right)^r \right)$$

(a form which leads more easily to asymptotic expression).

Both the distribution and the factorial cumulants of the classical sequential set-up discussed in the preceding paragraph are well known. It is known also that for k not too large

$$K_{[1]}(n) \doteq N \log \frac{N}{k + \frac{1}{2}}, \quad K_{[2]}(n) \doteq N^2 \left(\frac{\pi^2}{6} - \sum_{r=1}^k \frac{1}{r^2} \right).$$

When $k = 0$ the cumulants of
tend with increasing N to

$$x = \frac{n - \log N}{N}$$

$$K_1 = \gamma (\text{Euler's constant}), \quad K_r = \sum_{j=1}^{\infty} \frac{1}{j^r} \quad (r \geq 2).$$

This implies that in the limit

$$K_r = \left(-\frac{\partial}{\partial z} \right)^r \log \Gamma(z) \Big|_{z=1}.$$

These cumulants are the cumulants of a variable x whose cumulative probability distribution function is

$$\exp(-\exp(-x))$$

and this therefore is the limiting C.P.D.F.

The normal limit follows in the sequential case as in the fixed sample size, classical type occupancy problem.

Let $k = N(1-p)$ and allow N to increase without limit. The m th factorial cumulant of n can be written

$$\begin{aligned} K_{[m]} &= (-1)^{m-1} (N-k)(m-1)! + \sum_{r=0}^{N-k-1} \frac{(V/N)^m (m-1)!}{(1-V/N)^m} \\ &= (m-1)! \sum_{j=1}^m {}^m C_j (-1)^{m+j} \sum_{r=0}^{N-k-1} \left(1 - \frac{V}{N} \right)^j \end{aligned}$$

Let

$$T_j = \sum_{r=0}^{N-k-1} \left(1 - \frac{V}{N} \right)^{-j},$$

$$\frac{T_1}{N} = \sum_{r=0}^{N-k-1} \left(\frac{1}{N-V} \right) \rightarrow \log \frac{N}{k} = \log \frac{1}{q}$$

if $q = 1-p$.

$$\frac{T_j}{N} \rightarrow \int_0^p \frac{dx}{(1-x)^j} = -\frac{1}{j-1} \left(1 - \frac{1}{q^{j-1}} \right) \quad (\text{for } j \geq 2).$$

It follows that

$$K_{[m]} \doteq (m-1)! N \left\{ \left(\log \frac{1}{q} \right) m (-1)^{m-1} + \sum_{j=2}^m \frac{{}^m C_j (-1)^{m+j+1}}{j-1} \left(1 - \frac{1}{q^{j-1}} \right) \right\}$$

and

$$K_1 \doteq N \log \frac{1}{q}$$

$$K_2 \doteq -N \left(\log \frac{1}{q} \right) + \frac{pN}{q}.$$

As in the classical type problem this implies that

$$\frac{n - N \log(1/q)}{\sqrt{N\{(p/q) - \log(1/q)\}^{\frac{1}{2}}}}$$

is, asymptotically, a unit normal variable as $N \rightarrow \infty$.

ROMANOVSKY'S CASE WITH RESTRICTED COMPARTMENT CAPACITY

In this case, as we have described in our previous paper (Barton & David, 1959*a*), it is supposed that the capacity of each compartment is restricted to r balls. The probability distribution of n is, for k compartments, to be left empty,

$$\begin{aligned} p(n) &= \frac{r(k+1)}{Nr-n+1} \frac{{}^N C_{k+1}}{{}^{Nr} C_{n-1}} \sum_{j=0}^{N-k-1-\{(n-1)/r\}} (-1)^j {}^{N-k-1} C_j {}^{r(N-k-1-j)} C_{n-1} \\ &= \frac{{}^{N-1} C_k}{{}^{Nr-1} C_{n-1}} \sum_{j=0}^{N-k-1-\{(n-1)/r\}} (-1)^j {}^{N-k-1} C_j {}^{r(N-k-1-j)} C_{n-1}, \end{aligned}$$

a distribution which we believe is new. For the moments we proceed on similar, but not identical, lines to those for the classical sequential case so far considered. Suppose we enter the process at the point where t of the cells have been occupied, by S_t balls, and that n_t balls have to be dropped between the occupation of the t th and the $(t+1)$ st cells. We have, conditional on S_t (with conventionally $S_0 = 0$), that

$$p(n_t) = \frac{(rt - S_t)^{(n_t-1)}}{(Nr - S_t)^{(n_t)}} r(N-t) = - \frac{(rt - S_t)^{(n_t)}}{(Nr - S_t)^{(n_t)}} + \frac{(rt - S_t)^{(n_t-1)}}{(Nr - S_t)^{(n_t-1)}}.$$

It follows, still conditional on S_t , that

$$\mathcal{E} \left(\frac{(Nr - S_t - n_t + 1)^{[m]}}{(Nr - S_t + 1)^{[m]}} \right) = \frac{r(N-t)}{r(N-t) + m},$$

where

$$A^{[m]} = (A + m - 1)^{(m)}$$

for any number A and integer m . For the overall expectation we have

$$\mathcal{E} (Nr - S_{t+1} + 1)^{[m]} = \frac{N^{(t+1)} r^{t+1} (Nr + 1)^{[m]}}{P_{m,t+1}},$$

where

$$P_{m,t+1} = \prod_{j=0}^t (m + r(N-j)).$$

Reduction of this expectation gives us for the factorial advancing moment of S_t ,

$$\mathcal{E}(S_t^{[m]}) = (Nr + 1)^{[m]} N^{(t)} r^t \sum_{i=0}^m \frac{{}^m C_i (-1)^i}{P_{i,t}}.$$

n , the number of balls required to achieve just $N-k$ occupied boxes, will be equal to S_{N-k} . We have

$$\mathcal{E}(S_{N-k}) = \mathcal{E}(n) = (Nr + 1) \left[1 - \frac{N!}{\{N + (1/r)\}!} \frac{\{k + (1/r)\}!}{k!} \right]$$

$$\text{and } \mathcal{E}(n(n+1)) = (Nr + 1)(Nr + 2) \left[1 - \frac{2N!}{\{N + (1/r)\}!} \frac{\{k + (1/r)\}!}{k!} + \frac{N!}{\{N + (2/r)\}!} \frac{\{k + (2/r)\}!}{k!} \right],$$

where we use the notation $x! = \Gamma(x+1)$ for any $x > 0$. The higher factorial moments follow similarly.

APPARENT CONTAGION WITHIN A COMPARTMENT; A GENERALIZED PÓLYA SCHEME

Following our previous paper (1959*a*) if a is a measure of the contagion within a cell then the probability that $(k-1)$ cells out of N will be left empty when $(n-1)$ balls are dropped randomly is

$$\frac{N C_{k+1}}{a N + n - 2 C_{n-1}} \sum_{j=0}^{N-k-1} (-1)^j N^{-k-1} C_j^{a(N-k-1-j)+n-2} C_{n-1}.$$

As we remarked there, this is a generalization of Pólya's scheme. The probability that one further ball drops in an empty compartment is

$$\frac{a(k+1)}{a N + n - 1},$$

so that the probability distribution of n , the number of balls required to occupy $(N-k)$ compartments, is

$$p(n) = \frac{N^{-1} C_k}{a N + n - 1 C_{n-1}} \sum_{j=0}^{N-k-1} (-1)^j N^{-k-1} C_j^{a(N-k-1-j)+n-2} C_{n-1}.$$

The moments follow by arguments similar to those which we have used for the case immediately preceding. Alternatively, we note that formally we may write

$$a = -r, \quad (-A)^{(m)} = (-1)^m A^{[m]}.$$

Thus

$$\mathcal{E} \left(\frac{(Na + S_t + n_t - 1)^{(m)}}{(Na + S_t - 1)^{(m)}} \right) = \frac{(N-t)a}{a(N-t) - m}$$

conditional on S_t . The overall expectation is

$$\mathcal{E}(Na + S_{t+1} - 1)^{(m)} = \frac{N^{(t+1)} a^{t+1} (Na - 1)^{(m)}}{P_{m,t+1}},$$

where

$$P_{m,t+1} = \prod_{j=0}^t (a(N-j) - m).$$

Generally

$$\mathcal{E}(S_t^{(m)}) = (Na - 1)^{(m)} N^{(t)} a^t \sum_{i=0}^m \frac{m C_i (-1)^{i+m}}{P_{i,t}}.$$

GENERALIZED SEQUENTIAL OCCUPANCY: UNEQUAL PROBABILITIES IN EACH BOX

We consider here a multinomial type of sequential occupancy where there is no restriction as to box size but the probabilities for each box are not the same. Assume as before N boxes with probabilities $p_1, p_2, \dots, p_N$ of the balls falling in them. The balls are thrown independently until $(N-k)$ boxes have been occupied.

(i) n fixed and k , the number of zeros, a random variable

For this, the non-sequential case, we define a set of random variables α such that

$$\alpha_i = 1 \quad \text{if the } i\text{th box is empty after } n \text{ balls are dropped,} \\ = 0 \quad \text{if not.}$$

Then the number of empty boxes is

$$k = \sum_{i=1}^N \alpha_i$$

and

$$\mathcal{E}(k^{(m)}) = m! \sum_{i_1 < i_2 < \dots < i_m} (1 - p_{i_1} - \dots - p_{i_m})^n \\ = m! S_m^{(n)} (\text{say}) \quad (m \geq 1)$$

with

$$S_N^{(n)} = 0.$$

The probability generating function of k is

$$\Pi(z) = 1 + \sum_{m=1}^{N-1} (z-1)^m S_m^{(n)}$$

(ii) k fixed and n a random variable

Let $P_k(n)$ denote the cumulative probability distribution function of n . We have

$$\begin{aligned} P_k(n) &= \text{probability of at most } k \text{ zeros after } n \text{ throws} \\ &= \text{coefficient of } z^k \text{ in } \pi(z)/(1-z). \end{aligned}$$

Write

$$\pi^*(y) = \sum p_k(n) y^n,$$

where $p_k(n)$ is the P.D.F. of n . We have

$$\begin{aligned} \sum P_k(n) y^n &= \frac{\pi^*(y)}{1-y} \\ &= \text{coefficient of } z^k \text{ in } \frac{1}{1-z} \left\{ \frac{1}{1-y} + \sum_{m=1}^{N-1} (z-1)^m R_m(y) \right\}, \end{aligned}$$

where

$$R_m(y) = \sum_{i_1 < i_2 < \dots < i_m} \frac{1}{1 - (1 - p_{i_1} - \dots - p_{i_m}) y}.$$

It follows that

$$\frac{\pi^*(y) - 1}{y - 1} = \text{coefficient of } z^k \text{ in } \sum_{m=1}^{N-1} (z-1)^{m-1} R_m(y).$$

If further we put $y = 1 + \theta$,

$$\frac{\pi^*(y) - 1}{y - 1} = \sum_{V=1}^{\infty} \frac{\mu_{[V]}}{V!} \theta^{V-1}$$

and therefore

$$\frac{\mu_{[V]}}{V!} = \text{coefficient of } z^k \text{ in } \sum_{m=1}^{N-1} (z-1)^{m-1} \sum_{i_1 < \dots < i_m} \frac{[1 - (p_{i_1} + \dots + p_{i_m})]^V}{(p_{i_1} + \dots + p_{i_m})^V}.$$

TWO APPLICATIONS OF THE PREVIOUS SECTION

(i) Assume a line of unit length is broken, divided randomly into N pieces of lengths $p_1, p_2, \dots, p_N$. If y_m is the sum of m of these, then

$$p(y_m) = \frac{(N-1)!}{(m-1)!(N-m-1)!} y^{m-1} (1-y)^{N-m-1}.$$

If these lengths $p_1, p_2, \dots, p_N$ are now the box probabilities then the V th factorial moment of n , the number of balls required to occupy $N-k$ boxes is given by $\mu_{[V]}/V!$ which is the coefficient of z^k in

$$\sum_m (z-1)^{m-1} N C_m \frac{(N-m)^{[V]}}{(m-1)^{(V-1)}} (N-1).$$

(ii) Let the box probabilities bear a linear relationship one to the other so that

$$p_j = 2j/(N(N+1)) \quad (j = 1, 2, \dots, N).$$

Let us consider the mean number of balls necessary to ensure that every box is occupied. We have generally

$$\mu_{[1]} = \sum_m (-1)^{m-1} \sum \frac{1}{p_{i_1} + \dots + p_{i_m}},$$

where the inner sum is over all possible sets of m different p 's. For our particular case

$$\begin{aligned} (N+1)C_2^{-1} \mu_{[1]} &= \sum_{m=1}^N (-1)^{m-1} \sum \frac{1}{i_1 + \dots + i_m} \\ &= \int_0^1 \frac{\prod_{i=1}^N (1-x^i) - 1}{x} dx \end{aligned}$$

which may be shown to tend to $\{(4\pi/\sqrt{3}) - 6\}$. Thus

$$\mu_{[1]} \sim N+1 C_2 \left(\frac{4\pi}{\sqrt{3}} - 6 \right).$$

This mean value is very much larger than that for the classical sequential occupancy scheme, where if

$$p_i = \frac{1}{N}$$

we have

$$\mu_{[1]} \sim N \log N.$$

Ever since the early paper of Stevens (1937) the distribution of the number of zeros in the classical (De Moivre) case has been used in practice. Recently the Romanovsky type of problem has come to the front as, for example, where Turner & Eadie (1957) discussed the distribution of cells in haemocytometer counts when no overlapping of cells is permitted. So far we have not seen any applications of the sequential occupancy distributions other than in slightly unreal problems such as those connected with cigarette cards. There is no doubt, however, that sequential distributions are implicit in any collectors' problem. For example, the compartments may be traps and the arrival of the insects (i) may be random, (ii) may be conditioned by whether there is room in the traps, or (iii) may be decided by the equivalent of a gregarious spirit among the insects. These three conditions will correspond respectively to the three cases we have discussed, namely, the classical, the Romanovsky and the generalized Pólya scheme. It is also possibly worth noting that the sequential aspect of the occupancy problem may be considered as the number of renewals in a special form of renewal process.

REFERENCES

- BARTON, D. E. & DAVID, F. N. (1959a). *J.R. Statist. Soc. B.* (In the Press.)
 BARTON, D. E. & DAVID, F. N. (1959b). *J.R. Statist. Soc. B.* (In the Press.)
 FELLER, W. (1957). *An Introduction to Probability Theory*. New York: John Wiley and Sons.
 STEVENS, W. L. (1937). *Ann. Eugen., Lond.*, **8**, 57-69.
 TURNER, M. & EADIE, G. (1957). *Biometrics*, **13**, 485-95.

The generalized mean differences of the binomial and Poisson distributions

BY T. A. RAMASUBBAN

London School of Economics and Political Science

1. INTRODUCTION

Gini's mean difference g , defined by

$$g = \frac{2}{n(n-1)} \sum_i \sum_{j \neq i} |x_i - x_j| \quad (1.1)$$

is one of the well-known and useful measures of dispersion. Unlike most of the quantities designed with a view to measuring the amount of scatter, g is independent of any central measure of location. But, owing to the mathematical complexities introduced by its definition, the study of the distribution of this statistic and in particular some of its sampling properties has not advanced quite as much as in the case of other scale measures, except perhaps when the variates follow a normal law.

For the normal population, the exact standard error of g had been first given by Nair (1936) by a rather complicated method. Later Lomnicki (1952) obtained the same by a simpler derivation. Kamat (1953) evaluated the third moment in an exact form and, from the considerations of the measure of skewness β_1 , conjectured that for large n the distribution of g may be that of a χ . Following Kamat, I had obtained (1956) approximate values for its fourth moment and showed that the measure of kurtosis β_2 calculated therefrom, taken together with the values of β_1 obtained by Kamat, seem to establish the closeness of the χ -approximation, at least for sample size $n > 10$. I have also attempted to obtain empirical distributions of g for small samples ($n < 10$), the results of which I propose to report at a later date.

The normal distribution can be looked upon as a limiting case of the binomial and Poisson distributions under certain conditions. It is therefore natural to investigate the sampling moments of g for these discrete distributions and to attempt fitting an appropriate distribution for g even if the exact distribution is not easily found. In order to facilitate this work, the first step is to derive expressions for the absolute mean differences Δ_r given by

$$\Delta_r = \sum_i \sum_{j \neq i} |i-j|^r p(i)p(j) \quad (1.2)$$

for a few of the values of r , where i and j take all the integral values permitted by the discrete distribution $p(t)$. In a recent article (1958), I have succeeded in obtaining values of Δ_1 for some of the discontinuous distributions; in particular, those for the binomial and the Poisson distributions are shown to be:

$$\Delta_1(\text{binomial}) = 2npq \left[\sum_{i=0}^{n-1} \binom{n-1}{i}^2 p^{2i} q^{2n-2i-2} + \sum_{i=1}^{n-1} \binom{n-1}{i} \binom{n-1}{i-1} p^{2i-1} q^{2n-2i-1} \right] \quad (1.3)$$

and

$$\Delta_1(\text{Poisson}) = 2\lambda e^{-2\lambda} [I_0(2\lambda) + I_1(2\lambda)], \quad (1.4)$$

where $I_n(x)$ is the n th order modified Bessel function of the first kind. In the present paper I shall evaluate for these two distributions the generalized expressions for Δ_r valid for integral values of r .

2. THE EVEN ORDER MEAN DIFFERENCES

By definition, the even order mean difference is given by

$$\Delta_{2r} = \sum_i \sum_j |i-j|^{2r} p(i) p(j) = \sum_i \sum_j (i-j)^{2r} p(i) p(j). \quad (2.1)$$

Let ξ be the arithmetic mean and μ_k be the k th moment about ξ . Then, since $\mu_1 = 0$,

$$\begin{aligned} \Delta_{2r} &= \sum_i \sum_j \{(i-\xi) - (j-\xi)\}^{2r} p(i) p(j) = \mu_{2r} + \binom{2r}{2} \mu_{2r-2} \mu_2 - \binom{2r}{3} \mu_{2r-3} \mu_3 + \dots + \binom{2r}{2r-2} \mu_2 \mu_{2r-2} + \mu_{2r} \\ &= 2\mu_{2r} + \sum_{j=2}^{2r-2} (-1)^j \binom{2r}{j} \mu_{2r-j} \mu_j. \end{aligned} \quad (2.2)$$

It is thus possible to express the mean difference of even order in terms of the moments of the distribution. In particular,

$$\left. \begin{aligned} \Delta_2 &= 2\mu_2, \\ \Delta_4 &= 2(\mu_4 + 3\mu_2^2), \\ \Delta_6 &= 2(\mu_6 + 15\mu_4\mu_2 - 10\mu_3^2), \text{ and so on} \end{aligned} \right\} \quad (2.3)$$

On substituting the various values for the moments of the binomial and the Poisson distributions in (2.2) or (2.3), the corresponding even order mean differences can be evaluated.

3. THE ODD ORDER MEAN DIFFERENCES FOR THE BINOMIAL DISTRIBUTION

The evaluation of the odd order mean differences for the binomial distribution calls for some interesting algebraic manipulations. We have

$$\Delta_{2r+1} = \sum_{i=0}^n \sum_{j=0}^n |i-j|^{2r+1} \binom{n}{i} \binom{n}{j} p^i q^{n-i} p^j q^{n-j} \quad (3.1)$$

and since the sum of $(i-j)^{2r+1}$ over the distribution vanishes, (3.1) becomes

$$\begin{aligned} \Delta_{2r+1} &= 2 \sum_{i=0}^n \sum_{j=0}^i (i-j)^{2r+1} \binom{n}{i} \binom{n}{j} p^i q^{n-i} p^j q^{n-j} \\ &= 2 \sum_{i=0}^n \binom{n}{i} p^i q^{n-i} \left\{ \sum_{j=0}^i (i-j)^{2r+1} \binom{n}{j} p^j q^{n-j} \right\} \\ &= 2 \sum_{i=0}^n \binom{n}{i} p^i q^{n-i} \left\{ i^{2r+1} \binom{n}{0} q^n + (i-1)^{2r+1} \binom{n}{1} p q^{n-1} + \dots + \binom{n}{i-1} p^{i-1} q^{n-i+1} \right\}. \end{aligned} \quad (3.3)$$

Expanding the sum in (3.3) and rearranging the terms to form a finite series in $p^s q^{2n-s}$ ($s = 1, 2, \dots, 2n-1$) we have

$$\begin{aligned} \Delta_{2r+1} &= 2 \left[\sum_{i=0}^{n-1} \left\{ \sum_{j=0}^i (2j+1)^{2r+1} \binom{n}{i-j} \binom{n}{i+j+1} \right\} p^{2i+1} q^{2n-2i-1} \right. \\ &\quad \left. + \sum_{i=1}^{n-1} \left\{ \sum_{j=1}^i (2j)^{2r+1} \binom{n}{i-j} \binom{n}{i+j} \right\} p^{2i} q^{2n-2i} \right]. \end{aligned} \quad (3.4)$$

The method of evaluation of the sums

$$\sum_{j=0}^i (2j+1)^{2r+1} \binom{n}{i-j} \binom{n}{i+j+1} \quad \text{and} \quad \sum_{j=1}^i (2j)^{2r+1} \binom{n}{i-j} \binom{n}{i+j}$$

is somewhat involved and lengthy. I shall therefore omit such of the algebraical details which may not impair its clarity.

Consider the sum $\sum_{j=1}^i (2j)^{2r+1} \binom{n}{i-j} \binom{n}{i+j}$.

Let

$$S_k = \sum_{j=1}^{i-k} \binom{n-k}{i-j-k} \binom{n-k}{i+j} t^{2j+k}$$

and

$$\sigma_{k,i} = \sum_{j=1}^{i-k} (2j+k)^i \binom{n-k}{i-j-k} \binom{n-k}{i+j} t^{2j+k} = D^i(S_k),$$

where the operator $D \equiv t(\partial/\partial t)$. The sum required is therefore $\sigma_{0,2r+1}|_{t=1} = D^{2r+1}S_0|_{t=1}$; and since the expanded form of $D^{2r+1}S_0$ has $DS_0 (= \sigma_{0,1})$ as a factor, it is more convenient to operate $2r$ times on $\sigma_{0,1}$ and then set $t = 1$. Now

$$\sigma_{0,1} = \sum_{j=1}^i (2j) \binom{n}{i-j} \binom{n}{i+j} t^{2j}. \quad (3.5)$$

Splitting $2j$ as $\{(i+j) - (i-j)\}$, using Vandermonde's theorem, and finally summing over j , (3.5) gives

$$\sigma_{0,1} = n \left[t^2 \binom{n-1}{i} \binom{n-1}{i-1} + \left(t - \frac{1}{t} \right) \sum_{j=1}^{i-1} \binom{n-1}{i-j-1} \binom{n-1}{i+j} t^{2j+1} \right]. \quad (3.6)$$

On putting $t = 1$ in (3.6), we have the well-known result

$$\sigma_{0,1}|_{t=1} = n \binom{n-1}{i} \binom{n-1}{i-1}.$$

Again,

$$\sigma_{0,2r+1}|_{t=1} = D^{2r}\sigma_{0,1}|_{t=1} = n \left[\binom{n-1}{i} \binom{n-1}{i-1} D^{2r}(t^2) + \sum_{j=1}^{i-1} \binom{n-1}{i-j-1} \binom{n-1}{i+j} D^{2r} \left\{ \left(t - \frac{1}{t} \right) t^{2j+1} \right\} \right]_{t=1}. \quad (3.7)$$

It can readily be seen that any even number of operations by D on $(t - t^{-1})$ vanishes after t is put equal to unity and that any odd number of operations yields a factor 2. Therefore, expanding the term $D^{2r}\{(t - t^{-1}) t^{2j+1}\}$ on the extreme right of (3.7), by Leibniz's theorem and letting $t = 1$, we have

$$\begin{aligned} D^{2r}\{(t - t^{-1}) t^{2j+1}\}|_{t=1} &= 2 \left[\left\{ \binom{2r}{1} D^{2r-1} + \binom{2r}{3} D^{2r-3} + \dots + \binom{2r}{r-1} D^{r+1} \right\} t^{2j+1} \right]_{t=1} \\ &= 2 \sum_{p_1=1}^r \binom{2r}{2p_1-1} (2j+1)^{2r-2p_1+1} \end{aligned} \quad (3.8)$$

(3.7) can therefore be written as

$$\sigma_{0,2r+1}|_{t=1} = n \binom{n-1}{i} \binom{n-1}{i-1} 2^{2r} + 2n \sum_{p_1=1}^r \binom{2r}{2p_1-1} D^{2r-2p_1+1} S_1 \Big|_{t=1}. \quad (3.9)$$

Now

$$DS_1 = \sum_{j=1}^{i-1} (2j+1) \binom{n-1}{i-j-1} \binom{n-1}{i+j} t^{2j+1}.$$

Splitting $(2j+1)$ as $\{(i+j) - (i-j-1)\}$ and summing as above we find,

$$DS_1 = (n-1) \left[\binom{n-2}{i} \binom{n-2}{i-2} t^3 + \left(t - \frac{1}{t} \right) \sum_{j=1}^{i-2} \binom{n-2}{i-j-2} \binom{n-2}{i+j} t^{2j+2} \right]. \quad (3.10)$$

Operating on (3.10), $(2r-2p_1)$ times exactly as before and letting $t = 1$, we may show that

$$D^{2r-2p_1+1} S_1 = (n-1) \binom{n-2}{i} \binom{n-2}{i-2} 3^{2r-2p_1+2} (n-1) \sum_{p_2=1}^{r-p_1} \binom{2r-2p_1}{2p_2-1} D^{2r-2p_1-2p_2+1} S_2 \Big|_{t=1}. \quad (3.11)$$

Substitution of (3.11) in (3.9) and simplification yields,

$$\begin{aligned} \sigma_{0,2r+1}|_{t=1} &= a_{r1} n^{(1)} \binom{n-1}{i} \binom{n-1}{i-1} + a_{r2} n^{(2)} \binom{n-2}{i} \binom{n-2}{i-2} \\ &\quad + 2^2 \cdot n^{(2)} \sum_{p_1=1}^r \sum_{p_2=1}^{r-p_1} \binom{2r}{2p_1-1} \binom{2r-2p_1}{2p_2-1} D^{2r-2p_1-2p_2+1} S_2 \Big|_{t=1}, \end{aligned} \quad (3.12)$$

where

$$a_{r1} = 2^{2r} \quad \text{and} \quad a_{r2} = \frac{2^{2r+1}}{3.4} (2^{2r+1} - 2)$$

and $n^{(s)} = n(n-1)(n-2)\dots(n-s+1)$.

Dealing with (3.12) in exactly the same manner as with (3.9) whose forms are identical, we finally have

$$\sigma_{0,2r+1} \Big|_{t=1}^{r+1} = \sum_{s=1}^{r+1} a_{rs} n^{(s)} \binom{n-s}{i} \binom{n-s}{i-s}, \quad (3.13)$$

$$\text{where} \quad a_{rs} = 2^{s-1} \sum_{p_1=1}^r \sum_{p_2=1}^{r-p_1} \dots \sum_{p_{s-1}=1}^{r-\sum_{h=1}^{s-2} p_h} \binom{2r}{2p_1-1} \binom{2r-2p_1}{2p_2-1} \dots \binom{2r-\sum_{h=1}^{s-2} 2p_h}{2p_{s-1}-1} (s+1)^{2r-\sum_{h=1}^{s-1} 2p_h}. \quad (3.14)$$

It will be evident from (3.14) that the series (3.13) must terminate at $s = r+1$ since, for values of s greater than this, the last binomial coefficient occurring in (3.14) will vanish, even for the minimum value of the p_h , namely unity.

The evaluation of a_{rs} in (3.14) is not as difficult as it appears to be. The successive summations, starting with respect to p_{s-1} , ultimately give

$$a_{rs} = \frac{2^{2r+1}}{(s+1)\dots(2s)} \sum_{m=0}^{s-1} (-1)^m \left\{ \binom{2s-1}{m} - \binom{2s-1}{m-1} \right\} (s-m)^{2r+1}. \quad (3.15)$$

The procedure being the same for effecting the other sum, $\sum_{j=0}^i (2j+1)^{2r+1} \binom{n}{i-j} \binom{n}{i+j+1}$, I shall simply state the final result:

$$\sum_{j=0}^i (2j+1)^{2r+1} \binom{n}{i-j} \binom{n}{i+j+1} = \sum_{s=1}^{r+1} a'_{rs} n^{(s)} \binom{n-s}{i} \binom{n-s}{i-s+1}, \quad (3.16)$$

$$\text{where} \quad a'_{r,s+1} = \frac{1}{(s+1)\dots(2s)} \sum_{m=0}^s (-1)^m \left\{ \binom{2s}{m} - \binom{2s}{m-1} \right\} (2s-2m+1)^{2r}. \quad (3.17)$$

It may be pointed out here that there appears to be more than one method of evaluating these sums and that each method provides series similar to (3.13) and (3.16). However, these different series are only variations of the ones given here with the coefficients a_{rs} and a'_{rs} modified. It also seems possible to obtain a few relations concerning these coefficients but they are of only algebraic interest. The only relation which I shall make use of at a later stage is

$$a_{r,r+1} = a'_{r,r+1} = 2^{2r} r! \quad (3.18)$$

On substituting (3.13) and (3.16) in (3.4), the mean difference becomes

$$\Delta_{2r+1} = 2 \left[\sum_{i=0}^{n-1} \left\{ \sum_{s=1}^{r+1} a'_{rs} n^{(s)} \binom{n-s}{i} \binom{n-s}{i-s+1} \right\} p^{2i+1} q^{2n-2i-1} + \sum_{i=1}^{n-1} \left\{ \sum_{s=1}^{r+1} a_{rs} n^{(s)} \binom{n-s}{i} \binom{n-s}{i-s} \right\} p^{2i} q^{2n-2i} \right]. \quad (3.19)$$

It can be verified that on putting $r = 0$ in (3.19), the value of Δ_1 given by (1.3) is readily obtained.

(3.19) can be modified to give an alternative form for Δ_{2r+1} in terms of the hypergeometric function ${}_2F_1(\alpha, \beta; \gamma; x)$, where

$$F(\alpha, \beta; \gamma; x) = 1 + \frac{\alpha\beta}{\gamma} x + \frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)} \frac{x^2}{2!} + \dots$$

Taking account of the fact that the limits for i in (3.19) are to be taken such that none of the binomial coefficients like $\binom{n-s}{s}$ vanish, we find

$$\Delta_{2r+1} = 2 \sum_{s=1}^{r+1} \left\{ a'_{rs} n^{(s)} p^s q^s \sum_{i=s-1}^{n-s} \binom{n-s}{i} \binom{n-s}{i-s+1} p^{2i-s+1} q^{2n-2i-s-1} + a_{rs} n^{(s)} p^s q^s \sum_{i=s}^{n-s} \binom{n-s}{i} \binom{n-s}{i-s} p^{2i-s} q^{2n-2i-s} \right\}. \quad (3.20)$$

It is then not difficult to show that the sums in the braces on the right of (3.20) are respectively equal to

$$\binom{n-s}{s-1} (pq)^{s-1} {}_2F_1(-n+2s-1, \tfrac{1}{2}(2s-1); 2s-1; 4pq)$$

and

$$\binom{n-s}{s} (pq)^s {}_2F_1(-n+2s, \tfrac{1}{2}(2s+1); 2s+1; 4pq),$$

from the fact that these sums are the coefficients of $1/t^{s-1}$ and $1/t^s$ in $\{(q+pt)(q+p/t)\}^{n-s}$, i.e. in

$$\left\{1 + \frac{(1-t)^2}{t} pq\right\}^{n-s}, \quad \text{since } q+p=1.$$

Thus,

$$\Delta_{2r+1} = 2 \sum_{s=1}^{r+1} \left\{ a'_{rs} n^{(s)} \binom{n-s}{s-1} p^{2s-1} q^{2s-1} {}_2F_1(-n+2s-1, \tfrac{1}{2}(2s-1); 2s-1; 4pq) \right. \\ \left. + a_{rs} n^{(s)} \binom{n-s}{s} p^{2s} q^{2s} {}_2F_1(-n+2s, \tfrac{1}{2}(2s+1); 2s+1; 4pq) \right\}. \quad (3.21)$$

The following tables which will be useful in writing down explicitly the higher odd mean differences from either (3.19) or (3.21) show the numerical values of the coefficients a'_{rs} and a_{rs} for $r=0(1)5$; $1 \leq s \leq r+1$.

Table 1. Values of a'_{rs}

$r \backslash s$	1	2	3	4	5	6
0	1	.	.	.	.	.
1	1	4	.	.	.	.
2	1	40	32	.	.	.
3	1	364	1,120	384	.	.
4	1	3,280	30,912	32,256	6,144	.
5	1	29,524	799,040	1,951,488	1,013,760	122,880

Table 2. Values of $a_{rs}/2^{2r}$

$r \backslash s$	1	2	3	4	5	6
0	1	.	.	.	.	.
1	1	1	.	.	.	.
2	1	5	2	6	.	.
3	1	21	28	180	24	.
4	1	85	294	3,762	1,320	120
5	1	341	2,816			

4. THE ODD ORDER MEAN DIFFERENCES FOR THE POISSON DISTRIBUTION

The odd order mean difference Δ_{2r+1} for the Poisson distribution is given by

$$\Delta_{2r+1} = e^{-2\lambda} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} |i-j|^{2r+1} \frac{\lambda^i \lambda^j}{i! j!} \\ = 2e^{-2\lambda} \sum_{i=0}^{\infty} \sum_{j=0}^i (i-j)^{2r+1} \frac{\lambda^i \lambda^j}{i! j!} \quad (4.1)$$

for the same reason as for the binomial distribution.

The sums in (4.1) may be obtained in a straightforward manner; but this becomes quite tedious for $r > 2$. I have therefore considered the alternative simpler approach which is to obtain the limit of (3.21) for the binomial distribution as $n \rightarrow \infty$, $p \rightarrow 0$ and $np \rightarrow \lambda$.

Since $\lim_{\substack{n \rightarrow \infty, p \rightarrow 0 \\ np \rightarrow \lambda}} {}_2F_1(-n + \alpha, \beta; \gamma; 4pq) = {}_1F_1(\beta; \gamma; -4\lambda)$, (3.21) under these limiting conditions becomes

$$\Delta_{2r+1} = 2 \sum_{s=1}^{r+1} \left\{ a'_{rs} \frac{\lambda^{2s-1}}{(2s-1)!} {}_1F_1\left(\frac{1}{2}(2s-1); 2s-1; -4\lambda\right) + a_{rs} \frac{\lambda^{2s}}{(2s)!} {}_1F_1\left(\frac{1}{2}(2s+1); 2s+1; -4\lambda\right) \right\}. \quad (4.2)$$

Also, by virtue of the relation

$$\frac{\lambda^\nu}{\Gamma(\nu+1)} {}_1F_1\left(\frac{1}{2}(2\nu+1); 2\nu+1; -4\lambda\right) = e^{-2\lambda} I_\nu(2\lambda), \quad (4.3)$$

where the function I_ν has been explained earlier, (4.2) becomes

$$\Delta_{2r+1} = 2e^{-2\lambda} \sum_{s=1}^{r+1} \lambda^s \{a'_{rs} I_{s-1}(2\lambda) + a_{rs} I_s(2\lambda)\}. \quad (4.4)$$

It will be seen that the value of Δ_1 given by (1.4) agrees with the above, on letting $r = 0$; and

$$\left. \begin{aligned} \Delta_3 &= 2\lambda e^{-2\lambda} \{(1+4\lambda) I_0 + 4\lambda I_1\}, \\ \Delta_5 &= 2\lambda e^{-2\lambda} \{(1+16\lambda+32\lambda^2) I_0 + 8\lambda(1+4\lambda) I_1\}, \end{aligned} \right\} \quad (4.5)$$

after successive reduction of I_ν ($\nu \geq 2$) to I_0 and I_1 by the recurrence relation

$$I_{\nu+1}(2\lambda) = I_{\nu-1}(2\lambda) - (\nu/\lambda) I_\nu(2\lambda). \quad (4.6)$$

In (4.5), I_ν should be taken to mean $I_\nu(2\lambda)$. These values have also been checked independently by evaluating them directly from (4.1).

It is of some interest to consider the standardized ratio $\Delta_{2r+1}/(\text{variance})^{r+\frac{1}{2}}$, namely,

$$D_{2r+1} = \frac{2e^{-2\lambda}}{\lambda^{r+\frac{1}{2}}} \sum_{s=1}^{r+1} \lambda^s \{a'_{rs} I_{s-1}(2\lambda) + a_{rs} I_s(2\lambda)\}. \quad (4.7)$$

A limiting form for (4.7) can be got for large λ by considering the asymptotic series for the modified Bessel function $I_\nu(2\lambda)$. This is given by

$$I_\nu(x) \simeq \frac{e^x}{\sqrt{2\pi x}} \left\{ 1 - \frac{4\nu^2 - 1^2}{1! 8x} + \frac{(4\nu^2 - 1^2)(4\nu^2 - 3^2)}{2! (8x)^2} - \dots \right\} + O\left(\frac{e^{-x}}{\sqrt{x}}\right). \quad (4.8)$$

This serves all practical purposes so long as ν is not of order $\sqrt{x}$ (Lehmer, 1943). Considering only the first term in (4.8), the limiting value of D_{2r+1} for large λ can be written as

$$D_{2r+1} \simeq \frac{2e^{-2\lambda}}{\lambda^{r+\frac{1}{2}}} \sum_{s=1}^{r+1} \lambda^s \frac{e^{2\lambda}}{\sqrt{2\pi} \sqrt{2\lambda}} (a'_{rs} + a_{rs}),$$

which on simplification and omission of terms of order λ^{-1} , λ^{-2} , etc. gives us

$$D_{2r+1} \simeq \frac{1}{\sqrt{\pi}} (a'_{r,r+1} + a_{r,r+1}). \quad (4.9)$$

Substituting for the numerator from (3.18),

$$D_{2r+1} \simeq \frac{2}{\sqrt{\pi}} \cdot 2^{2r} \cdot r!. \quad (4.10)$$

When $r = 0$, the limiting value of D_1 equals $2/\sqrt{\pi}$, as demonstrated in one of the tables in my recent article (1958).

5. CONCLUSION

I may mention that although explicit results have been obtained here, it seems possible to derive certain interesting recurrence relations connecting the mean differences of different orders by considering the absolute moments of the distributions of the variate differences. This work is being pursued at the moment and I think I shall soon be able to record the outcome of this approach.

Finally, I wish to express my indebtedness to Prof. M. G. Kendall, under whose supervision the above investigations were carried out, for his continued support and encouragement.

REFERENCES

- KAMAT, A. R. (1953). The third moment of Gini's mean difference. *Biometrika*, **40**, 451.
 LEHMER, D. H. (1943). Note on the computation of the Bessel function $I_n(x)$. *Math. Tables and Aids to Computation*, **1**, 133.
 LOMNICKI, Z. A. (1952). The standard error of Gini's mean difference. *Ann. Math. Statist.* **23**, 635.
 NAIR, U. S. (1936). Standard error of Gini's mean difference. *Biometrika*, **28**, 428.
 RAMASUBBAN, T. A. (1956). A χ -approximation to Gini's mean difference. *J. Indian Soc. Agric. Statist.* **8**, 116.
 RAMASUBBAN, T. A. (1958). The mean differences and the mean deviations of some discontinuous distributions. *Biometrika*, **45**, 549.

Determination of parameters in the Johnson system of probability distributions

By D. C. M. LESLIE

Royal Aircraft Establishment, Farnborough

The system of curves proposed by Johnson (1949) for the graduation of statistical distribution has been very successful in practice. The principal drawback is that it is not very easy to determine the four parameters defining the distribution from the observed moments. In this note, a method is suggested for calculating these parameters if the distribution under study is not far removed from normality. For a short summary of the Johnson system, the reader is referred to Draper (1952).

The Johnson system of curves is mainly used for fitting to empirical data, but it can also be used to graduate theoretical distributions. In the author's researches, the generating function of a distribution was obtained in closed form, but the inversion integral could not be evaluated. The first four moments could be found very easily, and this suggested the use of a system of frequency curves to fit the distribution approximately. It was found that the distribution was never very far from normality, and that it gave rise to β_1, β_2 points which always lay in the S_U region of the β_1, β_2 plane. Therefore, this note deals only with the fitting of curves of the S_U family to distributions which are not too non-normal.

The variable x has an S_U distribution if

$$z = \eta + \delta \sinh^{-1} \left(\frac{x - \xi}{\lambda} \right), \quad (1)$$

z being a unit normal variable.* The problem is to determine η, δ, ξ and λ given the first four moments μ_r of an observed distribution (this term including distributions which arise theoretically). From the opposite point of view, it is comparatively easy to calculate the μ_r given η, δ, ξ and λ . Formulae are given by Johnson (1949, p. 163), and from these formulae it follows that

$$\beta_1 = \omega(\omega - 1) \frac{\{\omega(\omega + 2) \sinh 3\Omega + 3 \sinh \Omega\}^2}{2(\omega \cosh 2\Omega + 1)^3} \quad (2)$$

$$\beta_2 = \omega(\omega - 1) \frac{\omega^2(\omega^3 + 3\omega^2 + 6\omega + 6) \cosh 4\Omega + 4\omega(\omega + 3) \cosh 2\Omega - 3(\omega - 1)}{2(\omega \cosh 2\Omega + 1)^2}, \quad (3)$$

and

$$\gamma_2 = \beta_2 - 3 = (\omega - 1) \frac{\omega^2(\omega^3 + 3\omega^2 + 6\omega + 6) \cosh 4\Omega + 4\omega(\omega + 3) \cosh 2\Omega - 3(\omega - 1)}{2(\omega \cosh 2\Omega + 1)^2}, \quad (4)$$

where

$$\omega = e^{1/\delta^2} \quad \text{and} \quad \Omega = \eta/\delta.$$

The problem is now reduced to that of determining ω and Ω from given values of β_1 and γ_2 ; the functional form of equations (2) and (3) makes it clear that the problem is rather formidable.

However, it can be solved completely if the distribution is symmetrical about the origin, since then $\beta_1 = 0$, but $\gamma_2 \neq 0$. This implies $\Omega = 0$, so that equation (3) reduces to

$$\gamma_2 = \frac{1}{2}(\omega^2 - 1)(\omega^2 + 3) \quad (5)$$

and

$$\omega = \{\sqrt{(4 + 2\gamma_2)} - 1\}^{\frac{1}{2}} = \{\sqrt{(2\beta_2 - 2)} - 1\}^{\frac{1}{2}}.$$

* The letter η has been used in place of Johnson's γ because of the use of γ_2 for the standardized fourth cumulant.

This determines δ directly, while $\Omega = 0$ implies that $\eta = 0$ also. The remaining parameters ξ and λ are determined from the first two moments of x . Johnson's formulae are for the moments of the normalized variable

$$y = \frac{x - \xi}{\lambda}$$

and transforming them, we find $\mu'_1 = \xi - \lambda\omega^{\frac{1}{2}} \sinh \Omega$, (6)

$$\mu_2 = \frac{1}{2}\lambda^2(\omega - 1)(\omega \cosh 2\Omega + 1). \quad (7)$$

Since the distribution is symmetric, $\mu'_1 = 0$, while it has been proved that $\Omega = 0$; thus $\xi = 0$ also, while

$$\lambda = \sqrt{\frac{2\mu_2}{\omega^2 - 1}}, \quad (8)$$

ω being given by (5).

When $\beta_1 \neq 0$, the problem does not seem to be soluble in terms of recognized functions. However, equations (2) and (3) show that if $\beta_1 = \gamma_2 = 0$ (i.e. if the distribution is normal) then $\omega = 1$. This suggests that if β_1 and γ_2 are small, an expansion of $\omega - 1$ in powers of β_1 and γ_2 may be useful; the existence of such an expansion follows from the analytical nature of the expressions (2) and (3). To simplify the work, we put

$$m = \omega \sinh^2 \Omega \quad (9)$$

which reduces (2) and (3) to the purely algebraic forms

$$\beta_1 = (\omega - 1) \frac{m[4(\omega + 2)m + 3(\omega + 1)^2]^2}{2(2m + \omega + 1)^3}, \quad (10)$$

$$\gamma_2 = \frac{(\omega - 1)P(\omega)}{2(2m + \omega + 1)^2}, \quad (11)$$

where

$$P(\omega) = 8(\omega^3 + 3\omega^2 + 6\omega + 6)m^2 + 8(\omega^4 + 3\omega^3 + 6\omega^2 + 7\omega + 3)m + (\omega^5 + 3\omega^4 + 6\omega^3 + 10\omega^2 + 9\omega + 3).$$

Expanding equations (10) and (11) about the point $\omega = 1$, we find

$$\beta_1 = \frac{9m}{m+1}(\omega - 1) + O(\omega - 1)^2, \quad (12)$$

$$\gamma_2 = \frac{4(4m+1)}{m+1}(\omega - 1) + O(\omega - 1)^2, \quad (13)$$

(12) and (13) may be inverted to give

$$\omega = 1 - \frac{1}{3}\beta_1 + \frac{1}{4}\gamma_2 + O(\beta_1^2, \text{etc.}), \quad (14)$$

$$m = \frac{4\beta_1}{9\gamma_2 - 16\beta_1} + O(\beta_1, \gamma_2). \quad (15)$$

Equation (14) in particular is attractive in its simplicity, and provided β_1 and γ_2 are small, this pair of equations enables η and δ to be determined very rapidly. It should be noted that β_1 and γ_2 must be known fairly accurately if good values are to be obtained for η and δ .

If β_1 and γ_2 are not small enough for (14) and (15) to give good results, adequate accuracy can sometimes be obtained by using more terms of these expansions. The process outlined above has, therefore, been extended to include terms of order $(\omega - 1)^2$, and it is then found that

$$\omega = 1 - \frac{1}{3}\beta_1 + \frac{1}{4}\gamma_2 - \frac{31}{32}\beta_1^2 + \frac{11}{12}\beta_1\gamma_2 - \frac{1}{16}\gamma_2^2 + O(\beta_1^3), \quad (16)$$

$$m = \frac{4\beta_1}{9\gamma_2 - 16\beta_1} + \frac{(27\gamma_2^2 - 64\beta_1\gamma_2 + 40\beta_1^2)\beta_1}{6(9\gamma_2 - 16\beta_1)^2} + O(\beta_1^2). \quad (17)$$

If these expansions do not give adequate accuracy, they will at least provide useful starting values for a process of solving equations (10) and (11) by successive approximations; a considerable amount of labour would be needed to take them to the third order. Once η and δ have been found, ξ and λ may be determined from equations (6) and (7), which should be used in the forms

$$\mu'_1 = \xi - \lambda m^{\frac{1}{2}}, \quad (18)$$

$$\mu_2 = \frac{1}{2}\lambda^2(2m + \omega + 1). \quad (19)$$

It should be mentioned that Draper (1952) has solved the problem of inverting equations (2) and (3) heuristically. He has drawn extensive abacs of β_1 and γ_2 as functions of ω and Ω , and has then sought to fit empirical curves to these abacs. For ω he finds a power series not dissimilar to equation (16); he gives

$$\omega = 1 - 0.312\beta_1 + 0.227\gamma_2 - 0.047\beta_1^2 + 0.069\beta_1\gamma_2 - 0.025\gamma_2^2 + O(\beta_1^3),$$

while (16) is equivalent to

$$\omega = 1 - 0.333\beta_1 + 0.250\gamma_2 - 0.096\beta_1^2 + 0.153\beta_1\gamma_2 - 0.063\gamma_2^2 + O(\beta_1^3).$$

The coefficients of the first-order terms agree quite well, and those of the second-order terms have at least got the same signs. There is no very obvious correspondence between equation (17) and the various formulae given by Draper for Ω .

Dr Johnson has suggested a simple but instructive test of these formulae. A pair of values of ω and Ω is selected, and the corresponding values of β_1 and γ_2 calculated from equations (2) and (3); ω and Ω can then be estimated using the formulae given above, and the estimated values compared with the exact ones. Values obtained from equations (14) and (15) will be denoted by (ω_1, Ω_1) (first-order estimates) while those obtained from equations (16) and (17) will be denoted by (ω_2, Ω_2) (second-order estimates).

The table below shows the sort of results which are found when $(\omega - 1)$ is small

$\omega = 1.1 \}$	$\beta_1 = 0.0103 \}$	$\omega_1 = 1.111 \}$	$\omega_2 = 1.0983 \}$
$\Omega = 0.1 \}$	$\gamma_2 = 0.4560 \}$	$\Omega_1 = 0.097 \}$	$\Omega_2 = 0.1001 \}$
$\omega = 1.1 \}$	$\beta_1 = 0.2182 \}$	$\omega_1 = 1.112 \}$	$\omega_2 = 1.0979 \}$
$\Omega = 0.5 \}$	$\gamma_2 = 0.7386 \}$	$\Omega_1 = 0.480 \}$	$\Omega_2 = 0.5011 \}$
$\omega = 1.1 \}$	$\beta_1 = 0.5758 \}$	$\omega_1 = 1.115 \}$	$\omega_2 = 1.0969 \}$
$\Omega = 1.0 \}$	$\gamma_2 = 1.2268 \}$	$\Omega_1 = 0.914 \}$	$\Omega_2 = 0.9985 \}$
$\omega = 1.3 \}$	$\beta_1 = 0.0400 \}$	$\omega_1 = 1.405 \}$	$\omega_2 = 1.2401 \}$
$\Omega = 0.1 \}$	$\gamma_2 = 1.6744 \}$	$\Omega_1 = 0.089 \}$	$\Omega_2 = 0.1037 \}$

As long as γ_2 is not too large, the second-order formulae seem to give good results. Equation (5) shows that a series expansion in powers of γ_2 must diverge if γ_2 is greater than 2, and this implies that the method suggested above will only work well if $\gamma_2 < 1$. There must also be some limit on β_1 . Equation (5) is, however, valid for all values of γ_2 as long as $\beta_1 = 0$. This suggests that an expansion of the form

$$\omega = f_0(\gamma_2) + \beta_1 f_1(\gamma_2) + \beta_1^2 f_2(\gamma_2) + \dots$$

might be very useful. It is already known that

$$f_0(\gamma_2) = \{\sqrt{(4 + 2\gamma_2)} - 1\}^{\frac{1}{2}}$$

and the next term should not be too difficult to calculate.

The author is indebted to Dr G. M. Jenkins for drawing his attention to the problem and for giving much helpful advice on how to solve it, and to Dr N. L. Johnson for suggesting the test of accuracy given above. Acknowledgment is made to the Controller, H.M. Stationery Office, for permission to publish this note.

REFERENCES

- DRAPER, J. (1952). Properties of distributions resulting from certain simple transformations of the normal distribution. *Biometrika*, **39**, 290-301.
 JOHNSON, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, **36**, 149-76.

Note on a problem of estimation

By D. E. LLOYD

Ordnance Board, Ministry of Supply

INTRODUCTION

1. An estimation problem has arisen in which a random variable R from an unknown distribution with frequency function $p(r)$ has a probability of being observable which is a function of r denoted by $F(r)$. Thus the information available from an experiment consists of a set of observed values $r_1, r_2, \dots, r_m$,

together with the number $(n-m)$ of unobserved values. Truncated distributions are a special case which is frequently met in practice. However, in this note a case with a continuous function for $F(r)$ is considered.

2. In the examples considered, the information is used to estimate a parameter of the distribution function $p(r)$, $F(r)$ being assumed known. However, a similar method could be used to estimate parameters of $F(r)$. Estimates based on all the available information are compared with those based on $r_1, r_2, \dots, r_m$ only, and with those based on the ratio m/n .

3. The chance of an observed value r is $p(r) F(r)$. Thus, if a single value of R is chosen at random, the chance that it will be observable is

$$P = \int p(r) F(r) dr \quad (1)$$

and the chance that it is not observable is $(1-P)$, the integral being taken over the whole range of r .

4. If a sample of size n is taken from the population, the probability of the observed values being $r_1, r_2, \dots, r_m$ and there being $(n-m)$ unobserved values is

$${}^nC_m \prod_{i=1}^m \{p(r_i) F(r_i) dr_i\} (1-P)^{n-m}.$$

The log likelihood function is

$$L = \log {}^nC_m + \sum_{i=1}^m \log p(r_i) + \sum_{i=1}^m \log F(r_i) + (n-m) \log (1-P). \quad (2)$$

If θ is the parameter (or one of the parameters) of $p(r)$ to be estimated

$$\frac{\partial L}{\partial \theta} = \sum_{i=1}^m \frac{\partial}{\partial \theta} (\log p(r_i)) - \frac{n-m}{1-P} \frac{\partial P}{\partial \theta} = 0 \quad (3)$$

is the maximum likelihood equation for θ .

ESTIMATE BASED SOLELY ON NUMBER OF OBSERVED VALUES

5. In a sample of size n , the probability of obtaining exactly m observed values is

$${}^nC_m P^m (1-P)^{n-m}.$$

The log likelihood function $L = m \log P + (n-m) \log (1-P)$ so that the maximum likelihood equation is

$$\frac{\partial L}{\partial \theta} = \frac{m}{P} \frac{\partial P}{\partial \theta} - \frac{n-m}{1-P} \frac{\partial P}{\partial \theta} = 0,$$

therefore

gives the required estimate.

$$P(\hat{\theta}) = \frac{m}{n} \quad (4)$$

ESTIMATE BASED SOLELY ON $r_1, r_2, \dots, r_m$

6. Given that m realizations of R are recorded, the probability that the values $r_1, r_2, \dots, r_m$ will result is

$$\frac{\Pr \{r_1, r_2, \dots, r_m \text{ and } (n-m) \text{ unrecorded values}\}}{\Pr \{m \text{ recorded and } (n-m) \text{ unrecorded values}\}},$$

$$= \frac{\prod_{i=1}^m \{p(r_i) F(r_i) dr_i\} (1-P)^{n-m}}{P^m (1-P)^{n-m}},$$

therefore

$$L = \sum_{i=1}^m \log p(r_i) + \sum_{i=1}^m \log F(r_i) - m \log P.$$

Thus

$$\frac{\partial L}{\partial \theta} = \sum_{i=1}^m \frac{\partial}{\partial \theta} \log p(r_i) - \frac{m}{P} \frac{\partial P}{\partial \theta} = 0 \quad (5)$$

gives the required estimate.

Miscellanea

SPECIFIC FUNCTION FOR $p(r)$.

7. Now, put

$$p(r) = \frac{\alpha^{\frac{1}{2}\nu}}{\Gamma(\frac{1}{2}\nu)} r^{\frac{1}{2}\nu-1} e^{-\alpha r} \quad (0 \leq r < \infty).$$

This represents a χ^2 distribution with ν degrees of freedom and unknown dispersion.Then, $\log p(r) = \frac{1}{2}\nu \log \alpha + (\frac{1}{2}\nu - 1) \log r - \alpha r - \log \Gamma(\frac{1}{2}\nu)$,

hence

$$\frac{\partial}{\partial \alpha} \log p(r) = \frac{\nu}{2\alpha} - r.$$

Substituting in equation (3), we find

$$\frac{m\nu}{2\alpha} - \sum_{i=1}^m r_i - \frac{n-m}{1-P} \frac{\partial P}{\partial \alpha} = 0. \quad (6)$$

Put

$$A = \frac{1}{m} \sum_{i=1}^m r_i, \quad B = \frac{n-m}{m}. \quad (7)$$

Then equation (6) becomes

$$A - \frac{\nu}{2\alpha} + \frac{B}{1-P} \frac{\partial P}{\partial \alpha} = 0. \quad (8)$$

Thus A and B are jointly sufficient statistics for the estimation of α .

Equations (4) and (5) give

$$P(\alpha) = \frac{m}{n} \quad (9)$$

and

$$A - \frac{\nu}{2\alpha} + \frac{1}{P} \frac{\partial P}{\partial \alpha} = 0$$

for the other estimators of α .SPECIFIC FUNCTION FOR $F(r)$ 8. The function $F(r)$ must be chosen, as a result of prior information, to describe the conditions of the experiment. When truncation is assumed a function is chosen in the form

$$F(r) = 1 \quad (r \leq r_0), \\ = 0 \quad (r > r_0).$$

However, problems may arise in which, as a result of previous experiments, it seems more reasonable to use a continuous function for $F(r)$. We consider the case where

$$F(r) = \exp(-r/\lambda),$$

where λ is a constant, the value of which, it is assumed, is known.

9. For the particular functions chosen

$$P = \int_0^\infty F(r) p(r) dr \\ = \frac{\alpha^{\frac{1}{2}\nu}}{\Gamma(\frac{1}{2}\nu)} \int_0^\infty r^{\frac{1}{2}\nu-1} e^{-\{\alpha+(1/\lambda)\}r} dr \quad (11)$$

$$= \left(1 + \frac{1}{\alpha\lambda}\right)^{-\frac{1}{2}\nu}. \quad (12)$$

Hence

$$\frac{1}{1-P} \frac{\partial P}{\partial \alpha} = \frac{\nu}{2\alpha(\alpha\lambda+1)} \left\{ \left(1 + \frac{1}{\alpha\lambda}\right)^{\frac{1}{2}\nu} - 1 \right\} \quad (13)$$

and

$$\frac{1}{P} \frac{\partial P}{\partial \alpha} = \frac{\nu}{2\alpha(\alpha\lambda+1)}.$$

Equations (12) and (13) can be used to substitute for P in equations (8) and (10), respectively.

THE CASE $\nu = 2$

10. This is the case where r is the squared radial distance of a point distributed in two dimensions in circular normal form.

Equations (11), (12) and (13) give

$$P = \left(1 + \frac{1}{\alpha\lambda}\right)^{-1}, \quad (14)$$

$$\frac{1}{1-P} \frac{\partial P}{\partial \alpha} = \frac{1}{\alpha} \left(1 + \frac{1}{\alpha\lambda}\right)^{-1} \quad (15)$$

and

$$\frac{1}{P} \frac{\partial P}{\partial \alpha} = \frac{1}{\alpha} (1 + \alpha\lambda)^{-1}. \quad (16)$$

Using equation (15) to substitute for P in equation (8) gives

$$A\alpha^2 + \left(\frac{A}{\lambda} - 1 + B\right)\alpha - \frac{1}{\lambda} = 0. \quad (17)$$

The positive root of this equation gives the maximum likelihood estimate of α , denoted by $\hat{\alpha}_{ML}$.

11. Using equation (14) to substitute for P in equation (9) gives

$$1 + \frac{1}{\alpha\lambda} = \frac{n}{m},$$

therefore

$$\alpha = \frac{1}{\lambda B}. \quad (18)$$

We denote this estimate of α by $\hat{\alpha}_B$

12. Using equation (16) to substitute into equation (10) gives

$$A - \frac{1}{\alpha} + \frac{1}{\alpha} (1 + \alpha\lambda)^{-1} = 0.$$

Hence

$$\alpha = \frac{\lambda - A}{\lambda A}. \quad (19)$$

We denote this estimate of α by $\hat{\alpha}_A$

VARIANCES AND EFFICIENCIES OF ESTIMATES

13. For large n , the asymptotic variances of the three estimates are found to be

$$\text{var}(\hat{\alpha}_{ML}) = \frac{\alpha^2}{n} (P - P^2 + P^3)^{-1}, \quad (20)$$

$$\text{var}(\hat{\alpha}_B) = \frac{\alpha^2}{n} (P - P^2)^{-1}, \quad (21)$$

$$\text{var}(\hat{\alpha}_A) = \frac{\alpha^2}{n} P^{-3}. \quad (22)$$

If we measure 'efficiency' in terms of inverse variance and regard α_{ML} as being 100% efficient, Fig. 1 shows how the large sample efficiencies of $\hat{\alpha}_A$ and $\hat{\alpha}_B$ vary with P . $\hat{\alpha}_A$ is more efficient when P is large, $\hat{\alpha}_B$ when P is small, as might be expected. The sum of the two efficiencies is always 100%.

OTHER VALUES OF ν

14. For a value of $\nu \neq 2$ the equation corresponding to equation (17) will not reduce to a quadratic equation in α . In these cases it is sometimes convenient to construct a chart plotting A/λ against B for fixed values of $\alpha\lambda$. From equations (8) and (12) it is evident that the graphs are all straight lines, and approximate values of $\hat{\alpha}_{ML}$ can easily be found using the chart.

15. In conclusion I wish to thank Mr D. J. Bishop for advice and encouragement.

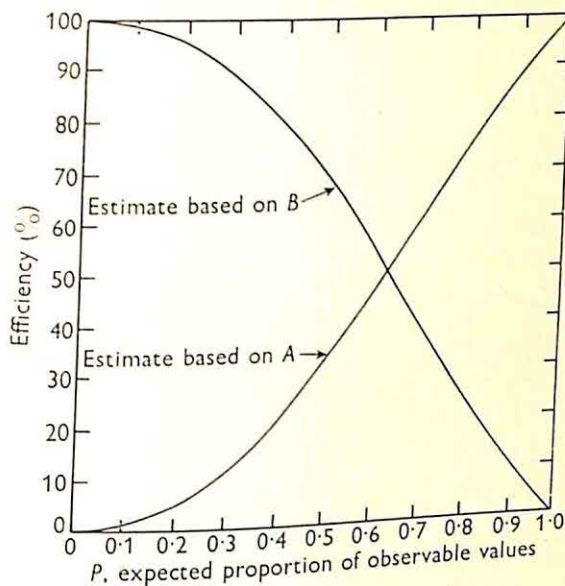


Fig. 1. Comparison of efficiencies.

On the non-central chi-square distribution

By MUNUSWAMY SANKARAN

Presidency College, Madras

1. Abdel-Aty (1954) has shown how the transformation

$$y = \left(\frac{\chi'^2}{f + \lambda} \right)^h = \left(\frac{\chi'^2}{r} \right)^h, \quad \text{where } r = f + \lambda \quad (1)$$

may be used to obtain an approximation to the distribution of non-central χ^2 , where f is the number of degrees of freedom, λ the non-central parameter and h an appropriate positive constant. To determine h , Aty examines the cumulants of y expressed in terms of cumulants of χ'^2 , as power series in r^{-s} . He finds 'after expansion that $h = \frac{1}{3}$ is the appropriate value to choose for minimizing the coefficient of r^{-2} in $\kappa_2(y)$ and reducing the higher cumulants, $\kappa_s(y)$ for $s > 2$, to a minimum'. He then develops two methods of approximation.

- (i) This assumes that $(\chi'^2/r)^{\frac{1}{3}}$ is normally distributed about $1 - \frac{2}{9}(1+b)/r$ with variance $\frac{2}{9}(1+b)/r$, where $b = \lambda/(f + \lambda)$.
- (ii) Alternatively, he suggests a 'closer approximation' which consists in applying the Cornish-Fisher (1937) expansion to the standardized deviate,

$$Y = \frac{y - \kappa_1(y)}{\sqrt{\{\kappa_2(y)\}}}, \quad (2)$$

where h is again taken as $\frac{1}{3}$.

2. In the present paper we shall use Aty's transformation (1) but instead of setting a fixed value to h , namely, $h = \frac{1}{3}$ here, we shall give h changing values so determined that the leading term in the third cumulant of y vanishes, thus following the procedure adopted by Wilson & Hilferty (1931). It follows that

$$h = 1 - \frac{k_1 k_3}{3k_2^2} = 1 - \frac{2(f + \lambda)(f + 3\lambda)}{(f + 2\lambda)^2}, \quad (3)$$

$$k_s = 2^{s-1}(s-1)!(f + s\lambda) \quad (4)$$

where

is the s th cumulant of χ'^2 . It will be seen from (3) that $\frac{1}{3} \leq h \leq \frac{1}{2}$. We have for the first four cumulants of y , for the value of h given by (3),

$$\left. \begin{aligned} \kappa_1(y) &= 1 + h(h-1) \frac{k_2}{2k_1^2} + h(h-1)(h-2)(1-3h) \frac{k_2^2}{8k_1^4} \\ &\quad + h(h-1)(h-2)(h-3) \frac{[2k_1^2 k_4 - (28-51h+11h^2)k_2^2]}{48k_1^6} + O\left(\frac{1}{r^4}\right), \\ \kappa_2(y) &= h^2 \frac{k_2}{k_1^2} + h^2(h-1)(1-3h) \frac{k_2^2}{2k_1^4} \\ &\quad + h^2(h-1) \frac{[(7h-11)k_1^2 k_4 - 2(h-2)(35h^2-84h+40)k_2^2]}{12k_1^6} + O\left(\frac{1}{r^4}\right), \\ \kappa_3(y) &= h^3(1-h) \frac{[2(23-49h+23h^2)k_2^2 - 3k_1^2 k_4]}{2k_1^6} + O\left(\frac{1}{r^4}\right), \\ \kappa_4(y) &= \frac{h^4}{3k_1^6} [3k_1 k_4 - 4(4-5h)k_2 k_3] + O\left(\frac{1}{r^4}\right), \\ \kappa_5(y) &= O\left(\frac{1}{r^4}\right) + \dots \end{aligned} \right\} \quad (5)$$

3. Two approximations will now be considered.

(iii) *First approximation.* We take y as normally distributed with the following mean and standard deviation:

$$\left. \begin{aligned} \mathcal{E}(y) &= 1 + h(h-1) \frac{k_2}{2k_1^2} - h(h-1)(2-h)(1-3h) \frac{k_2^2}{8k_1^4}, \\ \sigma(y) &= \frac{hk_2^{\frac{1}{2}}}{k_1} \left\{ 1 - \frac{(1-h)(1-3h)}{4k_1^2} k_2 \right\}. \end{aligned} \right\} \quad (6)$$

It should be noted that this approximation is similar to that given by Haldane (1937, p. 398) in his paper 'The approximate normalization of a class of frequency distributions'. We shall describe this as approximation H .

(iv) *Second approximation.** We shall in this case apply the Cornish-Fisher form of the Edgeworth expansion to the Y of equation (2), where h is as before given by (3) and the cumulants of y are obtained from (5), using the k_s of (4). It follows that considering terms up to order $r^{-\frac{1}{2}}$ †, the 100 α percentage point of the standardized deviate Y is given by

$$Y_\alpha = \xi_\alpha + \frac{1}{12}(\xi_\alpha^2 - 1)(1-h) \frac{[2(23-49h+23h^2)k_2^2 - 3k_1^2 k_4]}{k_1^3 k_2^{\frac{3}{2}}} + \frac{1}{72}(\xi_\alpha^3 - 3\xi_\alpha) \frac{[3k_1 k_4 - 4(4-5h)k_2 k_3]}{k_1 k_2^2},$$

where ξ_α is the 100 α percent point of a standardized normal deviate.

4. Some comparisons of approximations to the upper and lower 5% points are made in Table 1 below for the same f and λ values used by Aty. It is seen that at the upper 5% point the Edgeworth approximation (iv) of this paper, based on a varying h , is definitely more accurate than Aty's similar approximation (ii) with $h = \frac{1}{3}$. Further, the simpler approximation H , or (iii), is here somewhat better than Aty's Edgeworth approximation, and a great deal better than his simple normal approximation defined in (i) above (but not given in the table). At the lower 5% point both Edgeworth approximations are good, but Aty's is somewhat the more accurate. The H approximation is not so good here, and is not shown in the table.

5. It is difficult to draw any general conclusions because exact values for other percentage points are not available for comparison, and it is by no means certain that an approximation which is best at one percentage point will also be so at another. The present investigation has, however, shown that under some conditions approximations based on a fixed value of $h = \frac{1}{3}$ will not be the best to use.

* I owe this suggestion to Dr F. N. David.

† Strictly this includes also the term $\frac{1}{240}k_5(y)k_2(y)^{-\frac{1}{2}}(\xi_\alpha^4 - 6\xi_\alpha^2 + 3)$. It is felt that omission of this as well as terms of $O(r^{-4})$ in k_1, k_2, k_3 and k_4 will not affect the conclusions reached here. The inclusion of these terms will only strengthen the conclusions, but will also considerably increase the computations involved.

Table 1. Comparison of exact percentage points of the χ^2 distribution with the values obtained from various approximations

Upper 5% points						Lower 5% points		
f	λ	Exact	Aty's 'closer approx.'	Approx. H	Edgeworth expansion to the present approx.	Exact	Aty's 'closer approx.'	Edgeworth expansion to the present approx.
2	1	8.642	8.38	8.58	8.656	0.168	0.171	0.157
	4	14.641	14.62	14.63	14.661	0.646	0.643	0.654
	16	33.054	33.08	33.07	33.060	6.322	6.320	6.331
	25	45.308	45.33	45.31	45.317	12.080	12.077	12.085
4	1	11.707	11.67	11.69	11.681	0.909	0.908	0.918
	4	17.309	17.27	17.28	17.318	1.765	1.766	1.763
	16	35.427	35.44	35.43	35.435	7.884	7.885	7.893
	25	47.613	47.62	47.62	47.621	13.733	13.732	13.737
7	1	16.004	15.99	15.98	16.013	2.494	2.494	2.497
	4	21.228	21.21	21.20	21.242	3.664	3.665	3.668
	16	38.970	38.96	38.96	38.978	10.257	10.256	10.264
	25	51.061	51.06	51.06	51.068	16.226	16.227	16.231

My thanks are due to Dr F. N. David for her helpful criticism of an earlier version of the paper. I also had the benefit of helpful suggestions of Prof. E. S. Pearson in the preparation of the paper for publication. Finally, I wish to record my appreciation to some of my students for computational work in connexion with the table.

REFERENCES

- ABDEL-ATY, S. H. (1954). *Biometrika*, **41**, 538-40.
 CORNISH, E. A. & FISHER, R. A. (1937). *Rev. Inst. Int. Statist.* **5**, 307-20.
 HALDANE, J. B. S. (1937). *Biometrika*, **29**, 392-404.
 WILSON, E. B. & HILFERTY, M. M. (1931). *Proc. Nat. Acad. Sci., Wash.*, **17**, 684.

On the distribution of the largest of six roots of a matrix in multivariate analysis

By K. C. SREEDHARAN PILLAI* AND CELIA G. BANTEGUI†

The Statistical Center, University of the Philippines

1. INTRODUCTION

Roy (1945, 1953, 1957) has shown that tests of certain hypotheses and confidence interval estimation in multivariate situations can be based on those associated with the distribution of the extreme characteristic roots of a matrix whose elements are functions of the observations. The distribution of the extreme roots depends on the number of non-null roots, s , and on two other supplementary parameters, m and n , which are functions of the number of variates and, in some cases, the number of samples and the size of the samples.

* United Nations Senior Adviser in Mathematical Statistics and Visiting Professor of Statistics.

† Research Fellow, 1957-58, and presently Lecturer in Statistics.

Three different types of hypotheses are often posed for multivariate tests and the non-null characteristic roots of the matrix that arise in each test have a distribution, the form of which, under the null hypothesis, is the same for all the three tests. Roy (1939), Hsu (1939) and Fisher (1939) derived independently the distribution of the roots, $\theta_1, \dots, \theta_s$, which can be given in the form

$$f(\theta_1, \dots, \theta_s) = C(s, m, n) \prod_{i=1}^s \theta_i^m (1 - \theta_i)^n \prod_{i>j} (\theta_i - \theta_j) \quad (0 < \theta_1 \leq \dots \leq \theta_s < 1), \quad (1)$$

where

$$C(s, m, n) = \frac{\pi^{1/2} \prod_{i=1}^s \Gamma\left(\frac{2m+2n+s+i+2}{2}\right)}{\prod_{i=1}^s \Gamma\left(\frac{2m+i+1}{2}\right) \Gamma\left(\frac{2n+i+1}{2}\right) \Gamma\left(\frac{i}{2}\right)}, \quad (2)$$

and the parameters m and n are defined differently for the different types of hypotheses (Pillai, 1955).

In this paper, the cumulative distribution function of the largest of six roots has been studied with a view to tabulating the upper percentage points.

2. AN APPROXIMATION TO THE C.D.F. OF THE LARGEST OF SIX ROOTS OF A MATRIX

Pillai (1954) gave the explicit expression of the C.D.F. of θ_s for values of s ranging from two up to ten. Taking the case $s = 6$ and following his method (1954, 1956) an approximation to the C.D.F. of the largest of six roots, useful for computing the upper percentage points, has been obtained in the form:

$$\begin{aligned} \Pr(\theta_6 \leq x) = 1 - & \frac{\Gamma(2m+2n+12) \Gamma(m+3) \Gamma(n+3)}{4\Gamma(m+n+4) \Gamma(2m+6) \Gamma(2n+6)} x^{m+5} (1-x)^{n+1} \\ & + \frac{\Gamma(2m+2n+12) \Gamma(m+3) \Gamma(n+3)}{2(m+n+5) \Gamma(m+n+4) \Gamma(2m+5) \Gamma(2n+6)} x^{m+4} (1-x)^{n+1} \\ & - \frac{\Gamma(2m+2n+10) \Gamma(m+2) \Gamma(n+3) [3(m+n+5)(m+2) + (n+1)]}{2\Gamma(m+n+5) \Gamma(2m+4) \Gamma(2n+6)} x^{m+3} (1-x)^{n+1} \\ & + \frac{\Gamma(2m+2n+10) \Gamma(m+4) \Gamma(n+3)}{2(m+2) \Gamma(m+n+5) \Gamma(2m+3) \Gamma(2n+6)} x^{m+2} (1-x)^{n+1} \\ & - \frac{\Gamma(2m+2n+8) \Gamma(m+4) \Gamma(n+3)}{2\Gamma(m+n+4) \Gamma(2m+3) \Gamma(2n+6)} x^{m+1} (1-x)^{n+1}. \end{aligned} \quad (3)$$

For the details on the derivation of (3) the reader is referred to Bantegui (1958).

3. UPPER PERCENTAGE POINTS

Pillai (1954, 1956, 1957) using his approximation formulae, gave the upper 5 and 1 % points for the largest root for $s = 2$ and $s = 5$. These values were given for $m = 0(1)4$ and n varying from 5 to 1000 and are given to three or four significant figures to ensure sufficient accuracy for interpolation. Similarly, Sen (1957) gave tables of the upper 5 and 1 % points for 3 roots and Ventura (1957) for 4 roots. Pillai has published all these four tables, for $s = 2, 3, 4$ and 5 (1957).

Foster & Rees (1957) have tabulated the upper percentage points (80, 85, 90, 95 and 99) of the largest root for $s = 2$, $m = -0.5, 0(1)9$ and $n = 1(1)19(5)49, 59, 79$. Foster (1957, 1958) has further extended these tables for values of $s = 3$ and 4. The arguments they have used for tabulation are $\nu_1 = 2n + s + 1$ and $\nu_2 = 2m + s + 1$.

Bantegui (1958), also following the same approach as that of Pillai and using equation (3), obtained the percentage points for $s = 6$. These are presented in Tables 1 and 2. For computing the values of $\Pr(\theta_6 \leq x)$, a trial value was extrapolated from Pillai's tables (1957) of percentage points of the largest root for $s = 2, 3, 4$ and 5 for each of the $5 \times 16 = 80$ combinations of m and n ; the trial value and a few values around it were then examined and backward linear interpolation carried out using two values, one giving probability above and the other below the significance level.

4. ERROR OF APPROXIMATION

The error of approximation of the upper percentage points of the distribution is taken to be the difference between the approximate and exact probabilities. Pillai (1954), Sen (1957) and Ventura (1957) made such

Table 1. Upper 5% points of the largest root for $s = 6$

n	m				
	0	1	2	3	4
5	0.8246	0.8499	0.8685	0.8830	0.8945
10	.6552	.6917	.7206	.7442	.7639
15	.5371	.5758	.6077	.6346	.6577
20	.4535	.4912	.5231	.5505	.5746
25	.3918	.4276	.4583	.4852	.5091
30	0.3447	0.3782	0.4074	0.4332	0.4564
40	.2775	.3069	.3329	.3563	.3776
60	.1995	.2225	.2433	.2624	.2801
80	.1556	.1745	.1916	.2075	.2224
100	.1275	.1434	.1580	.1716	.1843
130	0.10036	0.11319	0.12504	0.13615	0.14666
160	.08272	.09348	.10388	.11284	.12175
200	.06702	.07586	.08409	.09186	.09926
300	.04545	.05156	.05728	.06281	.06790
500	.02765	.03143	.03498	.03835	.04160
1000	.01397	.01590	.01772	.01946	.02113

Table 2. Upper 1% points of the largest root for $s = 6$

n	m				
	0	1	2	3	4
5	0.8745	0.8929	0.9065	0.9169	0.9255
10	.7173	.7482	.7724	.7922	.8086
15	.5986	.6334	.6619	.6858	.7063
20	.5111	.5462	.5757	.6010	.6231
25	.4450	.4790	.5081	.5335	.5559
30	0.3936	0.4261	0.4542	0.4789	0.5011
40	.3194	.3484	.3739	.3969	.4177
60	.2315	.2548	.2757	.2948	.3125
80	.1814	.2006	.2181	.2342	.2493
100	.1491	.1654	.1803	.1942	.2072
130	0.11762	0.13091	0.14314	0.15457	0.16536
160	.09713	.10830	.11901	.12834	.13754
200	.07880	.08803	.09659	.10466	.11232
300	.05355	.05996	.06594	.07160	.07701
500	.03270	.03661	.04034	.04388	.04727
1000	.01651	.01855	.02046	.02229	.02405

studies for $s = 2, 3$ and 4 , respectively. A comparative study of the values of the error of approximation they have obtained indicates that;

(1) There is greater agreement between the probabilities for the approximate and exact cases in the upper 1 % points than in the 5 %.

(2) The difference between the approximate and the exact probabilities in the upper 5 % points occur in the fifth decimal place, which on rounding gives a difference of only one in the fourth decimal place.

(3) Taking any value of the parameter m constant, the error of approximation increases slowly as the other parameter is increased; such increase occurs only in the sixth decimal place or at most is unity in the fifth decimal place when rounding.

The error of approximation was similarly studied for $s = 6$. From Table 1, the percentage point for the parameter $m = 0, n = 500$ is 0.02765. At this percentage point, the numerical value of the approximate c.d.f. is 0.9500515, while the probability based on the exact formula is calculated to be 0.9500105. The error of approximation is then equal to 0.0000410 which occurs in the fifth decimal place. This result is consistent with that for lower values of s . It can therefore be concluded that the approximation formula (3) is sufficiently accurate for the computation of the upper 5% or lower percentage points of the c.d.f. of the largest of six roots.

It is to be noted that the error of approximation discussed above is not the same as the error in the percentage points. However, a comparison of the percentage points obtained, for instance, for $s = 3$ based on Pillai's approximate formula (Pillai, 1957) with Foster's table (1957) reveals that the approximation is sufficiently accurate to justify the retention of 4 and 5 decimal places for the 95 % level. In fact, the (1957) have also discussed this point for $s = 2$. Since the error of approximation obtained for $s = 6$ agrees with that computed for smaller values of s , the accuracy of the approximation for the present case is evident.

5. SUMMARY

In this paper the cumulative distribution function of the largest of six roots of a matrix in multivariate analysis has been studied. An approximation for obtaining upper percentage points (5 % or less) to the c.d.f. of the largest of six roots has been obtained following Pillai's method and numerical comparison was made of the probabilities obtained using the exact c.d.f. and the approximation to it. Since the comparison showed that the error of approximation was negligible, upper 5 and 1 % points were computed for small integral values of one parameter and values of a second parameter ranging up to 1000.

The authors wish to thank the Statistical Center, University of the Philippines for the facilities extended them during the preparation of this study.

REFERENCES

- BANTEGUI, C. G. (1958). 'On the distribution of the largest of six roots of a matrix in multivariate analysis.' Unpublished Thesis, The Statistical Center, University of the Philippines.
- FISHER, R. A. (1939). *Ann. Eugen., Lond.*, **9**, 238.
- FOSTER, F. G. (1957). *Biometrika*, **44**, 441.
- FOSTER, F. G. (1958). *Biometrika*, **45**, 492.
- FOSTER, F. G. & REES, D. H. (1957). *Biometrika*, **44**, 237.
- HSU, P. L. (1939). *Ann. Eugen., Lond.*, **9**, 250.
- PILLAI, K. C. S. (1954). 'On some distribution problems in multivariate analysis.' *Mimeo. Series* No. 88, Institute of Statistics, University of North Carolina.
- PILLAI, K. C. S. (1955). *Ann. Math. Statist.* **26**, 117.
- PILLAI, K. C. S. (1956). *Biometrika*, **43**, 122.
- PILLAI, K. C. S. (1957). *Concise Tables for Statisticians*. The Statistical Center, University of the Philippines.
- ROY, S. N. (1939). *Sankhyā*, **4**, 381.
- ROY, S. N. (1945). *Sankhyā*, **7**, 133.
- ROY, S. N. (1953). *Ann. Math. Statist.* **24**, 220.
- ROY, S. N. (1957). *Some Aspects of Multivariate Analysis*. New York: John Wiley and Sons, N.Y.
- SEN, P. (1957). 'On a multivariate test criterion and its applications.' Unpublished Thesis, The Statistical Center, University of the Philippines.
- VENTURA, S. R. (1957). 'On the extreme roots of a matrix in multivariate analysis and associated tests.' Unpublished Thesis, The Statistical Center, University of the Philippines.

The rank analogue of product-moment partial correlation and regression,
with application to manifold, ordered contingency tables†

By ROBERT H. SOMERS

Bureau of Applied Social Research, Columbia University

Daniels (1944) has demonstrated that the ordinary product-moment correlation, Spearman's rho, and Kendall's tau, are all special cases of a generalized correlation coefficient. This paper will demonstrate that a generalized *partial* correlation coefficient can be defined, as a simple extension of Daniels's theory. Kendall's partial tau will be seen to be a special case of this product-moment type partial coefficient. Thus, it is neither coincidental, nor remarkable (Kendall, 1955; Moran, 1950), that Kendall's partial tau is the same function of total taus that the product-moment partial is of total r . While the partial coefficient described here is of the first order, its form suggests a generalization to partial correlation of any order, which will be investigated at a later date. The theory will be seen to be immediately applicable to partial association in manifold, ordered, contingency tables, inasmuch as the partial tau here described is equally applicable to the case of ties among the rankings. As a further extension of product-moment theory to order statistics, the regression equations implied by Kendall's τ_a and τ_b will be described and illustrated, and the possibility of partial regression indicated.

No remarks are made here about sampling distributions. However, it is hoped that the generalized partial coefficient defined below will contribute to the difficult problem of the sample distribution of partial tau.

Daniels's generalized coefficient of correlation,

$$\Gamma = \frac{\sum a_{ij}b_{ij}}{\sqrt{\{\sum a_{ij}^2 \sum b_{ij}^2\}}},$$

is constructed in the following manner (cf. also Kendall, 1955, Ch. 2). When considering the correlation over n objects between two variables, P and Q , the objects are first numbered in arbitrary order, and all n^2 permuted pairs of objects are formed. Each pair is assigned a certain score, a_{ij} , for its value on variable P , and b_{ij} for variable Q . The ordinary product moment r becomes a special case of Γ when the assigned score, a_{ij} , is the difference in variate values of the i th and j th objects. Spearman's rho becomes a special case when the difference in ranking, rather than difference in variate value, is the score assigned to each pair. Finally, Kendall's tau is a special case of Γ when each pair is assigned a score of $\{+1/-1\}$ for pairs in which the j th member of the pair ranks {higher/lower} than the i th member, and 0 when the i th object is paired with itself, or when the i th and j th objects have the same rank.

This last case gives Kendall's τ_b . Since manifold, ordered, contingency tables may be considered rankings of objects with many ties, Kendall's τ_b is a useful coefficient of correlation for such data.

For convenience, we shall change Daniels's notation. We shall designate the scores as u_i for variable P , v_i for variable Q , and w_i for variable R , for $i = 1, 2, \dots, q$, where $q = n^2$. The manner of assigning scores is unchanged; we shall simply let Daniels' a_{ij} become $u_{n(i-1)+j}$, for example. Thus Daniels's generalized coefficient becomes

$$\Gamma_{PQ} = \frac{\sum u_i v_i}{\sqrt{\{\sum u_i^2 \sum v_i^2\}}} \quad (\text{for } i = 1, \dots, q).$$

In the following discussion we shall consider the u_i , v_i and w_i scores as components of vectors in q -dimensional Euclidian space, always taken relative to a fixed, normal, orthogonal basis. Thus, if we have n objects ranked on variables P , Q and R , we shall use the symbols $\mathbf{P}$, $\mathbf{Q}$ and $\mathbf{R}$ to represent vectors having $q = n^2$ components,

$$\mathbf{P} = (u_1, u_2, \dots, u_q), \quad \mathbf{Q} = (v_1, v_2, \dots, v_q), \quad \mathbf{R} = (w_1, w_2, \dots, w_q).$$

Further, as in ordinary vector analysis

$$|\mathbf{P}| = \sqrt{(u_1^2 + u_2^2 + \dots + u_q^2)},$$

† This article may be identified as Publication No. A-252 of the Bureau of Applied Social Research, Columbia University. I am indebted to my colleague, Herbert Menzel, whose analytical problem led to, and whose ingenuity aided in, the pursuit of these results. The preparation of this paper was facilitated by funds obtained from the Bureau of Applied Social Research.

and we shall designate the unit vectors

$$\alpha_1 = \frac{\mathbf{P}}{|\mathbf{P}|}, \quad \alpha_2 = \frac{\mathbf{Q}}{|\mathbf{Q}|}, \quad \alpha_3 = \frac{\mathbf{R}}{|\mathbf{R}|}.$$

The scalar ('inner' or 'dot') product of two vectors is, as usual, the scalar

$$(\mathbf{P} \cdot \mathbf{Q}) = u_1 v_1 + u_2 v_2 + \dots + u_q v_q.$$

At this point we diverge from the usual vector notation. We need to employ the notion of vector (or 'outer') product. But this is ordinarily defined only in three-dimensional space, and we operate in a space of far higher dimension. The generalization of the vector product to such a space ordinarily requires tensor notation (Kaplan, 1953, p. 161), but we shall avoid introducing new notation by defining a generalized vector product, which can be treated as a vector in a companion space, just as it is in 3-space;

$$(\mathbf{P} \times \mathbf{Q}) = (u_1 v_2 - u_2 v_1, u_1 v_3 - u_3 v_1, \dots, u_{q-1} v_q - u_q v_{q-1}),$$

i.e. a vector whose components are determinants of the 2×2 matrices of the form

$$\begin{vmatrix} u_i & u_j \\ v_i & v_j \end{vmatrix}, \quad \text{all } i < j, \text{ lexicographic ordering.}$$

This vector therefore has $\frac{1}{2}q(q-1)$ components, and is the vector perpendicular to the plane defined by the vectors $\mathbf{P}$ and $\mathbf{Q}$. (Note that $(\mathbf{P} \times \mathbf{Q}) = -(\mathbf{Q} \times \mathbf{P})$.)

Having established these definitions, we see immediately that

$$\tau_{PQ} = (\alpha_1 \cdot \alpha_2), \quad \tau_{PR} = (\alpha_1 \cdot \alpha_3) \quad \text{and} \quad \tau_{QR} = (\alpha_2 \cdot \alpha_3).$$

We shall show that

$$\tau_{QR.P} = \frac{(\alpha_1 \times \alpha_2) \cdot (\alpha_1 \times \alpha_3)}{|(\alpha_1 \times \alpha_2)| |(\alpha_1 \times \alpha_3)|} \quad (1)$$

is precisely the partial tau defined by Kendall as a phi coefficient[†] of a certain fourfold table (Kendall, 1955, p. 118), when the vectors α_i consist of scores appropriate to tau. With the appropriate scoring, this apparently becomes either a partial Spearman's rho, or the product-moment partial r.

From the analogous theory in 3-dimensional vector space, this form suggests that the partial can be treated as the cosine of the angle between the perpendiculars to the planes defined by the vectors α_1, α_2 and α_1, α_3 . Geometrically, this is seen to be equivalent to the representation of the ordinary product moment partial coefficient (cf. Kendall, 1948, p. 372) as the cosine of a different, but equal, angle—the angle between the two planes. In the ordinary representation of orthogonal projections on a subspace of one less dimension, the partial would be expressed as (cf. Birkhoff & MacLane (1949), p. 187)

$$\tau_{QR.P} = \frac{[\alpha_2 - (\alpha_1 \cdot \alpha_2) \alpha_1] \cdot [\alpha_3 - (\alpha_1 \cdot \alpha_3) \alpha_1]}{|[\alpha_2 - (\alpha_1 \cdot \alpha_2) \alpha_1]| |[\alpha_3 - (\alpha_1 \cdot \alpha_3) \alpha_1]|},$$

which is equivalent to the above form. However, the former has obvious advantages of representation, and immediately suggests the generalization to a partial of any order, since the vector product can be generalized to any number of vectors.

In order to show that (1) is the partial tau defined by Kendall, it is only necessary to observe that

$$(\alpha_1 \times \alpha_2) \cdot (\alpha_1 \times \alpha_3) = (\alpha_1 \cdot \alpha_1)(\alpha_2 \cdot \alpha_3) - (\alpha_1 \cdot \alpha_3)(\alpha_1 \cdot \alpha_2).$$

This is a well-known identity in 3-space, and will be demonstrated in the general case below. If this is true, then, since the α_i are unit vectors, $(\alpha_1 \cdot \alpha_1) = 1$. Moreover, $(\alpha_2 \cdot \alpha_3) = \tau_{QR}$, $(\alpha_1 \cdot \alpha_2) = \tau_{PQ}$ and $(\alpha_1 \cdot \alpha_3) = \tau_{PR}$. Thus the numerator is, as it should be, $\tau_{QR} - \tau_{PQ}\tau_{PR}$. In the denominator, we have

$$|(\alpha_1 \times \alpha_2)| = \sqrt{[(\alpha_1 \times \alpha_2) \cdot (\alpha_1 \times \alpha_2)]} = \sqrt{[(\alpha_1 \cdot \alpha_1)(\alpha_2 \cdot \alpha_2) - (\alpha_1 \cdot \alpha_2)(\alpha_1 \cdot \alpha_2)]} = \sqrt{(1 - \tau_{PQ}^2)};$$

and similarly for $|(\alpha_1 \times \alpha_3)|$. We have thus reduced the vector definition to Kendall's partial tau. Furthermore, the form of the partial immediately suggests the partial regression coefficients and partial covariance. It might be noted that the partial variance $(|\alpha_1 \times \alpha_2|)^2$ is the determinant $|A_2 A_2'|$, where A_2 is the $2 \times q$ matrix with rows α_1 and α_2 .

[†] The phi coefficient, $\frac{ad - bc}{\sqrt{\{(a+b)(a+c)(b+d)(c+d)\}}}$, it should be noted, is the product-moment correlation of a fourfold table.

We shall show that this equivalence between partial τ as the cosine of the angle between vectors, and partial τ as defined by Kendall, holds by enumerating the vector product. We have:

$$\begin{aligned}\alpha_1 &= (u'_1, u'_2, \dots, u'_q), \\ \alpha_2 &= (v'_1, v'_2, \dots, v'_q), \\ \alpha_3 &= (w'_1, w'_2, \dots, w'_q),\end{aligned}$$

i.e. unit vectors, where, for example, $u'_i = u_i/|P|$. Kendall's partial τ is

$$\tau_{QR.F} = \frac{\sum v'_i w'_i - \sum u'_i v'_i \sum u'_i w'_i}{\sqrt{[1 - (\sum u'_i v'_i)^2][1 - (\sum u'_i w'_i)^2]}}.$$

We will show that the numerator of this coefficient is equivalent to $(\alpha_1 \times \alpha_2) \cdot (\alpha_1 \times \alpha_3)$. Since we used the same expansion rule for vector products in establishing the denominator, this will suffice for establishing the equivalence.

By definition (and omitting primes on u'_i, v'_i and w'_i),

$$\begin{aligned}(\alpha_1 \times \alpha_2) \cdot (\alpha_1 \times \alpha_3) &= (u_1 v_2 - u_2 v_1)(u_1 w_2 - u_2 w_1) + \dots + (u_1 v_q - u_q v_1)(u_1 w_q - u_q w_1) \\ &\quad + (u_2 v_3 - u_3 v_2)(u_2 w_3 - u_3 w_2) + \dots + (u_{q-1} v_q - u_q v_{q-1})(u_{q-1} w_q - u_q w_{q-1}).\end{aligned}$$

This can be broken up into some positive terms with $u_i^2 v_j w_j$, and some negative terms of the form $u_i u_j (v_i w_j + v_j w_i)$, the former for $i \neq j$, the latter for $i < j$. These two groups of terms are reduced as follows:

$$\begin{aligned}\sum u_i^2 v_j w_j &= \sum_i u_i^2 \sum_j v_j w_j - \sum u_i^2 v_i w_i \quad (i \neq j, \text{ all } i, j = 1, \dots, q) \\ \sum_{i < j} u_i u_j (v_i w_j + v_j w_i) &= \frac{1}{2} [\sum_i u_i \sum_j u_j (v_i w_j + v_j w_i) - 2 \sum u_i^2 v_i w_i] \quad (\text{all } i, j).\end{aligned}$$

But, since these are unit vectors, $\sum u_i^2 = 1$. Thus

$$(\alpha_1 \times \alpha_2) \cdot (\alpha_1 \times \alpha_3) = \sum v_i w_i - \frac{1}{2} [\sum u_i u_j (v_i w_j + v_j w_i)] \quad (\text{all } i, j),$$

and we merely need to show that the negative term is $\sum u_i v_i \sum u_i w_i$. Since the matrix of terms $u_i u_j (v_i w_j + v_j w_i)$ is symmetric

$$\begin{aligned}\frac{1}{2} [\sum u_i u_j (v_i w_j + v_j w_i)] &= \sum_{i < j} u_i u_j (v_i w_j + v_j w_i) + \sum u_i^2 v_i w_i \quad (\text{all } i, j) \\ &= \sum_{i < j} u_i \sum u_j (v_i w_j) + \sum_{i < j} u_i \sum u_j (v_j w_i) + \sum u_i^2 v_i w_i = \sum u_i v_i \sum u_i w_i,\end{aligned}$$

and the equivalence is established.

A numerical example of the partial τ computed by means of the vector formulae will serve for illustration. For simplicity we shall avoid the redundancy involved in τ when computed from Daniels's generalized coefficient. As Kendall (1955, p. 20) notes, the Daniels's scoring rules give a numerator and denominator of τ twice the size of those given by the usual computing methods. Since notation increases the size of the vectors we have to consider, by adding redundant information, those consisting only of these vectors are large in any case, we have computed by means of reduced vectors, those consisting only of terms a_{ij} such that $i < j$.

Using the example of Kendall (1955, p. 117), and denoting the reduced vectors by the superscript, * we see that P, Q and R each have 15 components; the scores assigned to the 15 pairs as shown at the bottom of p. 117. Obviously $|P^*| = |Q^*| = |R^*| = \sqrt{15}$, and the α_i are unit vectors constructed in the usual fashion. Then $(\alpha_1 \times \alpha_2)^*, (\alpha_1 \times \alpha_3)^*$ and $(\alpha_2 \times \alpha_3)^*$ are vectors with $\frac{1}{2}m(m-1)$, or 105, components, where $m = \frac{1}{2}n(n-1)$, n being the number of objects ranked. Each component is a score assigned to each distinguishable pair of pairs, the score being $\pm 2/m$ or 0. In the tied case there will also be scores of $\pm 1/m$. In this example, we find

$$|(\alpha_1 \times \alpha_2)^*| = \sqrt{\{15^{-2}[4(25+19)]\}} = \sqrt{\frac{176}{15^2}},$$

$$|(\alpha_1 \times \alpha_3)^*| = \sqrt{\{15^{-2}[4(35+19)]\}} = \sqrt{\frac{216}{15^2}},$$

$$(\alpha_1 \times \alpha_2)^* \cdot (\alpha_1 \times \alpha_3)^* = \frac{4(6-15)}{15^2},$$

and

giving

$$\tau_{QR.P} = \frac{-36}{\sqrt{\{(216)(176)\}}} = -0.185,$$

in agreement with Kendall.

For those who work with manifold, ordered, contingency tables, it should be emphasized that these vector operations hold equally well for τ_b . Thus, when there are tied ranks

$$\tau_{bQR.P} = \frac{\tau_{bQR} - \tau_{bPQ}\tau_{bPR}}{\sqrt{\{(1 - \tau_{bPQ}^2)(1 - \tau_{bPR}^2)\}}}$$

is also the product-moment analogue of a partial correlation for contingency tables.†

As illustration of the meaning of Kendall's tau, it may be useful to consider the regression equations implied by the product-moment analogy. From our definition of $\tau_{QR} = (\alpha_2 \cdot \alpha_3) = \frac{(Q \cdot R)}{|Q| |R|}$, it is clear that the variance of scores on Q may be defined as $|Q|^2$. Thus, regression coefficients may be defined in the same way as in product-moment theory. In the case of total regression, for variables P and Q with no ties

$$\beta_1 = \beta_2 = \frac{(P \cdot Q)}{|Q|^2} = \frac{(P \cdot Q)}{|P|^2},$$

since $|P| = |Q| = n(n-1)$, where we have ranked n objects (without ties). The same example given by Kendall may be used to illustrate these coefficients. Here we have six objects ranked without ties, on variables P and Q . There are thus 36 possible pairs (considering all permutations) and six of them will be above, and following Daniels, these pairs give the following scatterplot.

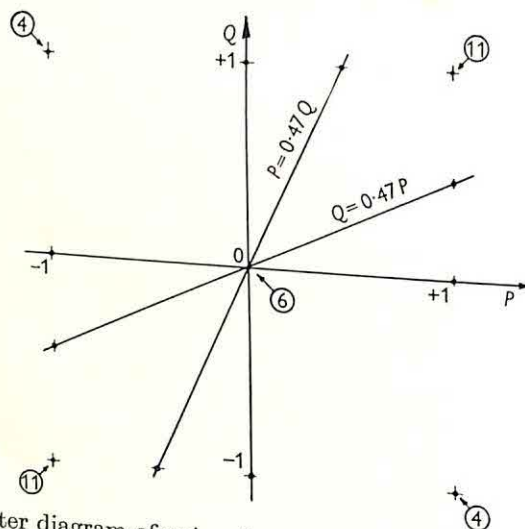


Fig. 1. Scatter diagram of pairs of observations implied by Kendall's τ_a .

(The circled figures are the number of permuted pairs given the indicated P and Q scores.)

$$\beta_1 = \beta_2 = \frac{(P \cdot Q)}{n(n-1)} = \frac{14}{30},$$

$\bar{P} = \bar{Q} = 0$, so the regression equations are as shown,
 $P = 0.47Q$ and $Q = 0.47P$,

$$\tau_{PQ} = \sqrt{(\beta_1 \beta_2)} = 7/15, \text{ in agreement with Kendall.}$$

In the case of tied rankings, say a contingency table as shown in Table 1 (from Lipset *et al.* (1951), p. 347) the scatterplot of pairs would be that given in Fig. 2.

† Such a partial coefficient for the tied case was used in an illustrative example, without justification, in a recent text; Sidney Siegel, *Nonparametric Statistics for the Behavioral Scientist*, p. 227. New York, 1956.

Table 1

Voting in accordance with predisposition	Level of knowledge				
		Low	Medium	High	Totals
	Yes	46	55	39	140
	No	9	35	38	82
	Totals	55	90	77	222

Since this is equivalent to ranking 222 objects, with many ties, there will be 49,284 permuted pairs entering into the scatterplot, in the following manner, where Q represents knowledge and R , vote.

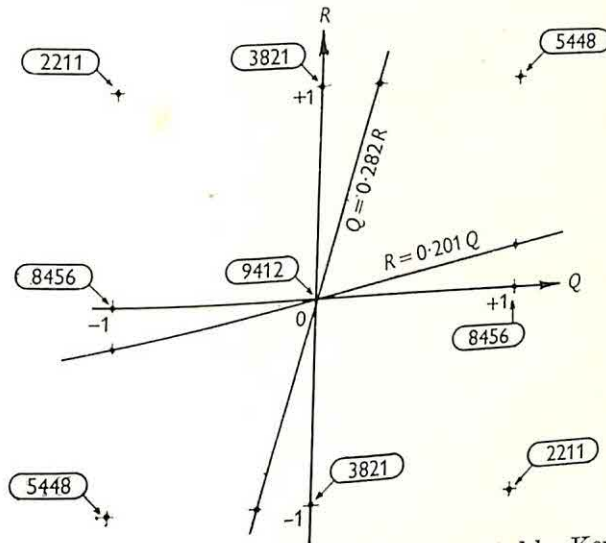


Fig. 2. Scatter diagram of pairs of observations implied by Kendall's τ_b .

(Again, the circled numbers are the number of permuted pairs at the given points.)

From the scatterplot, $\beta_1 = \frac{3237}{16115} = 0.201$, $\beta_2 = \frac{3237}{11480} = 0.282$

Again the means are 0, and the regression equations become,

$$Q = 0.201R, \quad R = 0.282Q.$$

The value of $\sqrt{(\beta_1\beta_2)} = \frac{3237}{\sqrt{\{(16115)(11480)\}}} = 0.238$. This, as we would expect, is exactly the value of τ_b for the contingency table given above. It will be noted that only in the case of ties among the rankings is it possible for the regression coefficients to be unequal. This is obvious from the redundancies in the scatterplot. It will be noticed that the first row of the scatterplot (Fig. 2) is the reverse of the last, and similarly for the columns. Again, this redundancy enters in because of the rule used in assigning scores; $a_{ij} = -a_{ji}$.

Further work is in progress on the order statistics analogues of product-moment statistics. On the basis of Daniels's initial conceptualization, it seems clear that a general partial and multiple correlation and regression theory for order statistics can be generated. Similarly, the correlation ratios, for the case of tied

ranks, seem to be equivalent to assymetric coefficients of association for a contingency table that does not have ordered variables, i.e. attaining their maxima when the observations in any row or column fall entirely within one cell.

Finally, we may suggest an alternative point of view to those who reject such product moment derivatives as ϕ on the grounds that they have no 'operational meaning' (Goodman & Kruskal, 1954). While such meaning for a coefficient is often desirable, there are other times when *theoretical power* assumes greater importance. Often both theoretical power and operational meaning cannot be attained. The present investigations have resulted from a need for a coefficient of association in manifold, ordered, contingency tables which would permit logical inferences to be made about operationally undefined intervening variables. Thus a coefficient was needed which would give information about a partial coefficient when only the total coefficients were known. Obviously, if a choice had to be made in this case, theoretical power could not be sacrificed for the sake of operational meaningfulness.

REFERENCES

- BIRKHOFF, G. & MACLANE, S. (1949). *A Survey of Modern Algebra*. New York: The Macmillan Company.
- DANIELS, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika*, **33**, 129-35.
- GOODMAN, L. & KRUSKAL, Wm. (1954). Measures of association for cross classification. *J. Amer. Statist. Assoc.* **49**, 732-64.
- KAPLAN, W. (1953). *Advanced Calculus*. Cambridge, Massachusetts: Addison Wesley.
- KENDALL, M. G. (1948). *The Advanced Theory of Statistics*, **1**. London: Charles Griffin and Co. Ltd.
- KENDALL, M. G. (1955). *Rank Correlation Methods* (2nd ed.). London: Charles Griffin and Co. Ltd.
- LIPSET, S. M., TROW, M. COLEMAN, J. (1956). *Union Democracy*. Glencoe Illinois: Free Press.
- MORAN, P. (1950). Recent developments in ranking theory. *J. Roy. Statist. Soc. Ser. B*, **12**, 153-62.

The busy period in relation to the queueing process GI/M/1

By B. W. CONOLLY

ABSTRACT. A treatment by difference equations is given of the probability distributions associated with length of the Busy Period in connexion with the simple queueing process with general input and random service. A numerical comparison is made in a few typical cases between the processes $M/M/1$ and $D/M/1$, the latter being of interest in connexion with appointment systems.

1. INTRODUCTION

Kendall (1953) suggested the notation $GI/G/N$ as a compact description of an N -server queueing process in which the inter-arrival intervals are independently and arbitrarily distributed, and the service intervals are also arbitrarily distributed. 'GI' refers to input, and 'G' to service.

The process which will be considered here has the general independent input, while the service has probability density function (P.D.F.) of the negative exponential, or 'Markovian', type, viz.

$$e^{-t/b}/b. \quad (1.1)$$

Service is supplied by one server, and the queue is dealt with in order of arrival. Inter-arrival intervals and service intervals are statistically independent.

A consequence of the negative exponential assumption for service is that the probability $b_n(t)$ of the completion of n consecutive service periods in an interval $(0, t)$, the $(n+1)$ th being in progress at the instant t , is given by

$$b_n(t) = e^{-t/b}(t/b)^n/n!. \quad (1.2)$$

Denoting the P.D.F. of inter-arrival intervals by

$$a(t) \quad (0 \leq t < \infty) \quad (1.3)$$

with mean a , we shall indicate by $c_n(t)$ the P.D.F. associated with an inter-arrival interval of length t during which time n consecutive service periods are completed, the $(n+1)$ th being in progress at the end of the interval. Thus,

$$c_n(t) = a(t) b_n(t). \quad (1.4)$$

Most of the following investigation will involve Laplace transforms. In general, we shall use the corresponding Greek letter for the transform of a function denoted by a Roman letter.

2. THE BUSY PERIOD

From the server's viewpoint the progress of a queueing system of the type considered here is a succession of alternating periods of idleness and full occupation. It is the latter which are called Busy Periods (often abbreviated to B.P.'s.), and the object of the investigation is to make probability statements about the lengths of B.P.'s.

The length in time can conveniently be treated in the manner of Bailey (1954). Essentially this consists in a study of the special process the starting point of which is the arrival and immediate reception of service of a customer, and which terminates when the server becomes idle once more. The crux of the present treatment is the simultaneous investigation of length, measured in time and number served, by an integro-difference equation technique of the type used in Conolly (1958). This is a versatile method, and the author has in essence used it also to study the B.P. in connexion with the system $GI/E_k/1$ (an extension of the present work), and with $E_k/G/1$. Earlier work by other writers is mentioned in Kendall (1951), and, more recently, there is a comprehensive study in connexion with the processes $M/G/1$ and $D/G/1$ by Takács (1955).

To be specific, in this paper we shall investigate the probability density functions $p_m(t)$ associated with the termination of a B.P. in the small time interval $(t, t+dt)$ after it began, a total of m customers having received service. Then

$$p_m = \int_0^\infty p_m(t) dt \quad (2.1)$$

is the probability that a total of m customers receive service during a B.P., whatever its length in time, and

$$p(t) = \sum_{m \geq 1} p_m(t) \quad (2.2)$$

is the P.D.F. associated with termination of the B.P. in $(t, t+dt)$, however many customers are served. In conditions favouring a steady state we shall expect to find that

$$\int_0^\infty p(t) dt = 1. \quad (2.3)$$

Otherwise the integral will be less than unity and a definite probability will exist of B.P.'s. of infinite duration.

3. BASIC THEORY

We introduce a P.D.F. $f_{mn}(t)$ associated with the following composite event:

- (i) a customer arrives at $t = 0$ and immediately begins to receive service;
- (ii) a further arrival takes place in the subsequent small interval $(t, t+dt)$;
- (iii) as a result of this arrival the total number waiting in the queue is n ($n \geq 1$);
- (iv) at the time t of this arrival the m th customer to be served in succession is at the counter ($m \geq 1$).

Then, for $n = 1$, we have

$$f_{m1}(t) = \int_0^t \sum_{r=1}^{m-1} f_{m-r,r}(s) c_r(t-s) ds, \quad (3.1)$$

for $m \geq 2$, together with

$$f_{11}(t) = c_0(t).$$

The explanation of (3.1) is that at the time of the previous arrival, in $(s, s+ds)$, say, any customer from the first up to the $(m-1)$ th could have been at the counter. If it was the $(m-r)$ th the previous arrival must have created a queue of r , of which the last member must be at the counter at t . The arrival at t is then the only one waiting.

Jumping directly to Laplace transforms we obtain

$$\phi_{11}(z) = \gamma_0(z), \quad (3.2)$$

and

$$\phi_{m1}(z) = \sum_{r=1}^{m-1} \phi_{m-r,r}(z) \gamma_r(z), \quad (3.3)$$

where $\gamma_r(z)$ is the Laplace transform of $c_r(t)$.

Generalizing (3.4) for $n \geq 2$, and omitting arguments, we obtain,

$$\phi_{mn} = \sum_{r=0}^{m-1} \phi_{m-rn-1+r} \gamma_r. \quad (3.5)$$

Writing

$$\Phi_n(y, z) = \sum_{r \geq 1} y^r \phi_{rn}(z)$$

for the generating function of the Laplace transforms ϕ_{mn} , we obtain by addition of (3.3) and (3.4) multiplied by appropriate powers of y the infinite difference equation

$$\Phi_1(y, z) = y\gamma_0(z) + \sum_{r \geq 1} y^r \gamma_r(z) \Phi_r(y, z); \quad (3.6)$$

and for $n \geq 2$,

$$\Phi_n(y, z) = \sum_{r \geq 0} y^r \gamma_r(z) \Phi_{n-1+r}(y, z). \quad (3.7)$$

Following the usual method for the solution of this type of difference equation we substitute, in (3.7), x^n for Φ_n . The equation thus becomes

$$x = \Gamma(xy, z), \quad (3.8)$$

where

$$\begin{aligned} \Gamma(y, z) &= \sum_{r \geq 0} y^r \gamma_r(z) \\ &= \int_0^\infty a(t) \exp[-t(z + (1-y)/b)] dt \\ &= \alpha(z + (1-y)/b). \end{aligned} \quad (3.9)$$

If $\xi_r(y, z)$ are the roots in x of (3.8) ($r \geq 1$),

$$\Phi_n(y, z) = \sum_r K_r(y, z) \xi_r^n(y, z), \quad (3.10)$$

the numbers K being independent of n .

Now it is obvious that for $|y| \leq 1$

$$|\Phi_n(y, z)| \leq |\Phi_n(1, z)|. \quad (3.11)$$

Also $\sum_n \Phi_n(1, z)$ is the Laplace transform of the probability of an arrival while the B.P. is in progress and is certainly finite while $R(z) > 0$. Hence for $R(z) > 0$ and $|y| \leq 1$ the summation in (3.10) can only include ξ_r with modulus less than unity. By applying Rouché's theorem for $R(z) > 0$ we find that (3.8) has only one root ξ that fits this description, and substituting $\Phi_n = K\xi^n$ for $n \geq 1$ into (3.6) we obtain $K = y$. Thus,

$$\Phi_n(y, z) = y\xi^n(y, z). \quad (3.12)$$

4. THE LENGTH OF A BUSY PERIOD

Our final objective is to obtain formulae connected with the P.D.F.'s, $p_m(t)$ (cf. § 2). We shall denote by $\pi_m(z)$ the Laplace transforms of these functions, and our objective will be realized by a formula for the generating function $\Pi(y, z)$ of the $\pi_m(z)$, i.e. for

$$\Pi(y, z) = \sum_{m \geq 1} y^m \pi_m(z). \quad (4.1)$$

Then $\Pi(y, 0)$ is the generating function of the probabilities $p_m(2.1)$ that a B.P. concludes after the service of the m th customer, while $\Pi(1, z)$ is the L.T. of the P.D.F. $p(t)$ (2.2) associated with the termination of a B.P. in $(t, t+dt)$.

We write

$$A(s) = \int_s^\infty a(t) dt, \quad (4.2)$$

and $d_n(t)$ will denote the L.T. of the d_n , and its generating function

$$\Delta(y, z) = \sum_{n \geq 1} y^n \delta_n(z) \quad (4.3)$$

can readily be seen to be given by

$$\Delta(y, z) = \sum_{n \geq 1} y^n \delta_n(z) \quad (4.4)$$

$$\begin{aligned} \Delta(y, z) &= \int_0^\infty e^{-(z+(1-y)/b)s} A(s) ds \\ &= b[(1 - \Gamma(y, z))/(bz + 1 - y)], \end{aligned} \quad (4.5)$$

since

$$\int_0^\infty e^{-zs} A(s) ds = (1 - \alpha(z))/z.$$

Now

$$p_1(t) = A(t) e^{-tb}/b, \quad (4.6)$$

since the probability that the B.P. terminates in $(t, t+dt)$ after the completion of service of the first customer is simply the joint probability that no one arrives in $(0, t)$ and that the service of the first customer is completed in $(t, t+dt)$. For $m \geq 2$ we have

$$p_m(t) = \frac{1}{b} \int_0^t \sum_{r=1}^{m-1} f_{m-r,r}(s) d_r(t-s) ds, \quad (4.7)$$

since, if, at the arrival before t , the $(m-r)$ th customer was at the counter, this same arrival must have created a queue of r ($1 \leq r \leq m-1$), and the whole of this queue plus the customer at the counter must be disposed of in the remaining time interval, the last in $(t, t+dt)$, during which time no further arrivals can take place. The associated L.T.'s are

$$\pi_1(z) = \delta_0(z)/b, \quad (4.8)$$

and, for $m \geq 2$,

$$\pi_m(z) = \frac{1}{b} \sum_{r=1}^{m-1} \phi_{m-r,r}(z) \delta_r(z). \quad (4.9)$$

By comparing coefficients of powers of y it will be seen that

$$\begin{aligned} \Pi(y, z) &= [y\delta_0(z) + \sum_{r \geq 1} y^r \delta_r(z) \Phi_r(y, z)]/b \\ &= y\Delta(\xi y, z)/b \\ &= y(1 - \xi)/(bz + 1 - \xi y). \end{aligned} \quad (4.10)$$

5. LIMITING BEHAVIOUR OF $\Pi(y, z)$

Although (4.10) has only been proved under the conditions $|y| \leq 1$ and $Re z > 0$, its validity can be extended to include $z = 0$ if $|y| < 1$, and, in circumstances to be investigated, if $y = 1$. In point of fact $\Pi(y, 0)$ and $\Pi(1, 0)$ are of direct practical importance; the former since it is the generating function of the probabilities p_m relating to the length of a B.P. in terms of customers served; the latter, since it is the probability of termination in finite time. Moreover, the value N of

$$\left[\frac{d\Pi(y, 0)}{dy} \right]_{y=1},$$

is the average number of customers served during a B.P.

The limiting behaviour of $\Pi(y, z)$ is required essentially for all values of the congestion index $\rho = b/a$. We commence with the study of $\Pi(1, z)$. By (4.10) this is

$$\Pi(1, z) = (1 - \xi)/(bz + 1 - \xi), \quad (5.1)$$

$$x = \Gamma(x, z) = \alpha(z + (1 - x)/b). \quad (5.2)$$

where ξ is the smallest root of

$$\text{If we can expand } \xi \text{ into a Taylor series in } z \text{ we obtain} \quad (5.3)$$

$$\xi(z) = \sum_{n \geq 0} k_n z^n, \quad (5.4)$$

where

$$k_n = \left[\frac{d^n \xi}{dz^n} \right]_{z=0} / n!. \quad (5.5)$$

Thus

$$\Pi(1, z) = \frac{1 - k_0 - k_1 z - 0(z)}{bz + 1 - k_0 - k_1 z - 0(z)}.$$

By the theory of Branching Processes it is known that:

- (i) When $\rho < 1$, $k_0 < 1$. It is evident that as $z \rightarrow 0$, $\Pi(1, z) \rightarrow 1$, exactly as would be expected in steady state conditions.
- (ii) When $\rho > 1$, $k_0 = 1$, and as $z \rightarrow 0$, $\Pi(1, z) \rightarrow k_1/(k_1 - b) = \rho^{-1}$. It follows that a non-zero probability exists of infinitely long B.P.'s.
- (iii) When $\rho = 1$, $\xi = 1$ is a double root of (5.2) with $z = 0$. The expansion of $\xi(z)$ proceeds in non-integral powers, but its leading term is unity and again $\Pi(1, z) \rightarrow 1$ as $z \rightarrow 0$.

Formulae for N can be obtained from

$$\Pi(y, 0) = (1 - \xi)y/(1 - \xi y), \quad (5.6)$$

where ξ is the smallest root in x of

$$x = \alpha[(1 - xy)/b].$$

Assuming it to be legitimate, we expand ξ into a series of ascending powers of $(1 - y)$, viz.

$$\xi = \sum_{n \geq 0} h_n (1 - y)^n,$$

where

$$h_n = \frac{(-)^n}{n!} \left[\frac{d^n \xi}{dy^n} \right]_{y=1}.$$

Differentiating (5.6) and letting $y \rightarrow 1$ we obtain for $\rho < 1$

$$N = (1 - h_0)^{-1}. \quad (5.7)$$

The limiting process is a little more complicated when $\rho > 1$. On the understanding that we refer only to B.P.'s. of finite duration we obtain

$$N = (h_2 - h_1) \rho / h_1^2. \quad (5.8)$$

6. THE SYSTEM $D/M/1$

It is well known that regularizing the input (which is the significance of D) has the effect of reducing customers' average waiting time as compared with a random input, and any appointment system is a recognition of the fact. We can now study its effect on a B.P. If a is the interval between successive arrivals, we have

$$\alpha(z) = e^{-az}; \quad (6.1)$$

then in connexion with numbers of customers served our interest centres on the smallest root ξ of

$$\xi = e^{-(1-\xi v)\rho}. \quad (6.2)$$

This can be expanded into the series

$$\xi = \frac{\rho}{y} \sum_{n \geq 1} \frac{(ny/\rho e^{1/\rho})^n}{n(n!)}, \quad (6.3)$$

for instance, by assuming a Taylor expansion and calculating the coefficients by direct differentiation of (6.2). The result is quoted in Takács (1955).

If (6.3) be substituted into (5.6), then a little rearrangement enables the probabilities p_m concerning the termination of a B.P. after m customers have been served to be picked out as the coefficients of the powers of y . For instance,

$$p_1 = 1 - e^{-1/\rho}, \quad (6.4)$$

$$p_2 = e^{-1/\rho} - e^{-2/\rho}(1 + 1/\rho), \quad (6.5)$$

$$p_3 = e^{-2/\rho}(1 + 1/\rho) - e^{-3/\rho}(1 + 2/\rho + 3/2\rho^2), \quad (6.6)$$

$$p_4 = e^{-3/\rho}(1 + 2/\rho + 3/2\rho^2) - e^{-4/\rho}(1 + 3/\rho + 4/\rho^2 + 8/3\rho^3), \quad (6.7)$$

$$p_5 = e^{-4/\rho}(1 + 3/\rho + 4/\rho^2 + 8/3\rho^3) - e^{-5/\rho}(1 + 4/\rho + 15/2\rho^2 + 25/3\rho^3 + 125/24\rho^4). \quad (6.8)$$

These formulae can be used for all values of ρ .

For the evaluation of (5.8) we have

$$h_1 = -(\rho - 1)^{-1}, \quad (6.9)$$

$$h_2 = (3\rho - 2)/2(\rho - 1)^3. \quad (6.10)$$

and

In the following Table 1 we give values of the first five p_m for a selection of values of ρ in the two cases $M/M/1$ and $D/M/1$, together with

$$\sum_{m=1}^5 p_m, \quad \sum_{m=1}^{\infty} p_m \quad \text{and} \quad N.$$

It will be noticed that regularizing arrivals has the effect of increasing the probabilities of shorter B.P.'s. However, in the single instance considered with $\rho > 1$, the average number served during a finite B.P. is greater for the regular than for the random input.

Table 1. Numerical quantities associated with the busy period for $M/M/1$ and $D/M/1$.

ρ	p_1	p_2	p_3	p_4	p_5	$\sum_{i=1}^5 p_m$	$\sum_{i=1}^{\infty} p_m$	N
<i>M/M/1</i>								
0.25	0.8000	0.1280	0.0410	0.0164	0.0073	0.9927	1.0000	1.3333
0.50	.6667	.1482	.0658	.0366	.0228	.9401	1.0000	2.0000
1.00	.5000	.1250	.0625	.0391	.0273	.7539	1.0000	∞
1.50	.4000	.0960	.0461	.0276	.0186	.5883	.6667	3.000
<i>D/M/1</i>								
0.25	0.9817	0.0166	0.0015	0.0002	0.0000	1.0000	1.0000	1.0174
0.50	.8647	.0803	.0277	.0124	.0063	.9914	1.0000	1.2550
1.00	.6321	.0973	.0465	.0287	.0199	.8245	1.0000	∞
1.50	.4866	.0741	.0334	.0190	.0124	.6255	0.6667	4.5000

REFERENCES

- BAILEY, N. T. J. (1954). A continuous time treatment of a simple queue using generating functions. *J.R. Statist. Soc. B*, **16**, 288-91.
- CONOLLY, B. W. (1958). A difference equation technique applied to the simple queue with arbitrary arrival interval distribution. *J.R. Statist. Soc. B*, **20**, 168-75.
- KENDALL, D. G. (1951). Some problems in the theory of queues. *J.R. Statist. Soc. B*, **13**, 151-85.
- KENDALL, D. G. (1953). Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. *Ann. Math. Statist.* **24**, 338-54.
- TAKÁCS, L. (1955). Investigation of waiting time problems by reduction to Markov processes. *Acta Math. Acad. Sci. Hung.* **6**, 101-28.

**The design of factorial experiments: a survey of some schemes requiring
not more than 256 treatment combinations**

BY R. G. MITTON AND F. R. MORGAN

British Leather Manufacturers' Research Association, Milton Park, Egham, Surrey, England

INTRODUCTION

Factorial experiments involving factors at two levels only are particularly simple to design and analyse and are widely used in consequence. Because 4- or 8-level factors can be regarded as formally equivalent to two or three factors at two levels, these can be incorporated also without trouble. The method for preparing designs for such experiments is straightforward (Fisher, 1926, 1935; Yates, 1937), but if they involve many factors the search for suitable designs may require some time. A number of designs for factorials with factors at 2, 4, and 8 levels are given in this paper in Tables 1-10. The schemes tabulated require up to 256 treatment combinations, and use is made of fractional replication (Finney, 1945), but not partial confounding. It is assumed that second and higher-order interactions can be used for the estimate of error, and for confounding (Fisher, 1935; Finney, 1947), but main effects are unconfounded and, where possible, first-order interactions also. In the schemes with fractional replication, no main effect or first-order interaction has as alias an interaction of lower order than the second.

The tables have been made as complete as possible (subject to the restrictions stated), but it may well be that some of the designs can be improved by modifications which have not been found during our search. A few of the schemes are essentially similar to those given by other workers; e.g. Yates, 1937; Brownlee, Kelley & Loraine, 1948; Cochran & Cox, 1950; and U.S. Department of Commerce, 1957.

EXPLANATION OF THE TABLES

The standard notation developed by Fisher (1935) is used, capital letters denoting the main effects and interactions, and small letters the treatment levels. Where a scheme involves 8-level factors these are written first and allocated the first letters of the alphabet (*ABC* for the first 8-level factor, *DEF* for the second). Four-level factors are written next, and each is allocated two of the first available letters; the remaining letters (with *I* and *O* omitted) represent 2-level factors.

Each scheme is given a reference number consisting of five digits. The first three of these give the number of 8-level, 4-level and 2-level factors, respectively. Fractions of a replicate are indicated by the fourth digit *m*, and the number of blocks by the fifth digit *n*; the scheme is a $(\frac{1}{2})^m$ replicate in 2^n blocks. Where necessary, brackets are used in the scheme number; e.g. 10(11)63 indicates a scheme for one 8-level factor and eleven 2-level factors, and is a $\frac{1}{64}$ replicate in 8 blocks.

All the schemes in any one table have the same number of (equivalent) 2-level factors, and the table gives the scheme number, the generators of the defining contrasts and of the interactions confounded, the generators of the principal block, the generators of the multipliers which operate on the members of the principal block to give the remaining blocks, and the degrees of freedom available for estimating error. In several of the schemes one or more first-order interactions are confounded with blocks; these are tabulated.

The statistical analysis of the data from an experiment with fractional replication is carried out by first neglecting certain of the factors so that a complete replicate of those remaining is obtained. The contrasts are then obtained for this remainder in the usual way (Finney, 1946; Yates, 1937). All the schemes given are such that the last letters can be omitted for this purpose; thus, scheme 10(11)63 has 14 letters of which 6 are to be omitted, leaving *ABCDEFGH*.

We wish to thank the Director and Council of the British Leather Manufacturers' Research Association for permission to publish this paper.

REFERENCES

- BROWNLEE, K. A., KELLY, B. K. & LORAIN, P. K. (1948). Fractional replication arrangements for factorial experiments. *Biometrika*, **35**, 268.
- COCHRAN, W. G. & COX, G. M. (1950). *Experimental Designs*. New York: Wiley and Sons.
- FINNEY, D. J. (1945). Fractional replication of factorial experiments. *Ann. Eugen., Lond.*, **12**, 291.
- FINNEY, D. J. (1946). Recent developments in the design of field experiments. III. Fractional replication. *J. Agric. Sci.* **36**, 184.
- FINNEY, D. J. (1947). Construction of confounding arrangements. *Emp. J. Exp. Agric.* **15**, 107.
- FISHER, R. A. (1926). The arrangement of field experiments. *J. Minist. Agric.* **33**, 503.
- FISHER, R. A. (1935). *The Design of Experiments*. Edinburgh: Oliver and Boyd.
- YATES, F. (1937). The design and analysis of factorial experiments. *Tech. Commun. Bur. Soil Sci., Harpenden*, no. 35.
- U.S. DEPARTMENT OF COMMERCE (1957). Fractional factorial experiment designs for factors at two levels. *Nat. Bur. Standards Appl. Math. Series*, no. 48.

Tables 1-10

Table 1. Designs of type 2⁵

Tables 1-10

Table 1. *Designs of type 2⁵*

Scheme number	Generators of defining contrasts	Generators of interactions confounded with blocks	Generators for principal block	Multipliers for other blocks	D.F. for error	1st-order interactions confounded
Full-replicate designs: 32 treatment combinations						
00502 01302	} —	ACE, BDE	abe, ac, bd	a, b {	13	—
					10	—
02101 10201	} —			ABCDE	ab, ac, ad, ae	a {
		6	—			

Table 2. Designs of type 2⁶

Table 2. *Designs of type 2⁶*

Full-replicate designs: 64 treatment combinations															
00603 01403 02203 03003	}	—	<i>ABCD, ABEF, ACE</i>	<i>acf, ade, bce</i>	<i>a, b, c</i> {	35 31 27 23	— — <i>ABCD</i> <i>ABCD, ABEF, CDEF</i>								
02202 03002 10302 11102						}	—	<i>ACDE, ABDF</i>	<i>ad, bf, ce, abc</i>	<i>a, b</i> {	30 24 26 19	— — — <i>ACDE</i>			
Half-replicate designs: 32 treatment combinations															
00613	}	<i>ABCDEF</i>	<i>ACE, BDE, BCF</i>	<i>abcd, abef</i>	<i>ab, ac, ae</i>	6	<i>AB, CD, EF</i>								
00612 01412						}	<i>ACD, BCF</i>	<i>ad, bf, abce</i>	<i>ab, ac</i> {	8 5	<i>CE</i> <i>CE</i>				
00611 01411										}	<i>ADE</i>	<i>ad, ae, bc, bf</i>	<i>ab</i> {	9 5	— —

Table 4. *Designs of type 2^s*

Scheme number	Generators of defining contrasts	Generators of interactions confounded with blocks	Generators for principal block	Multipliers for other blocks	D.F. for error	1st-order interactions confounded
Full-replicate designs: 256 treatment combinations						
00804 01604	}	—	ABCD, AB EF, ABGH, AC FH	a, b, c, e	{ 204 198	— —
02404 03204 04004						
02404 03204 04004	}	—	ACH, AEG, BFH, DFG	a, b, c, d	{ 191 183 174	— — —
10504 11304 12104						
20204 21004						
10504 11304 12104 20204 21004	}	—	BEH, BFG, ABDE, ACDF	a, b, c, e	{ 183 174 164 149 137	ABDE BFG, ABDE ABDE, ACDF, BCEF ABDE, ACDF, BCEF, EFGH, BCGH
11303 12103						
20202 21002						
11303 12103	}	—	BFH, BCDF, ABEFG	a, b, c	{ 181 170	— —
20202 21002						
20202 21002	}	—	ABDEG, ACDFH	a, b	{ 158 144	— —
Half-replicate designs: 128 treatment combinations						
00813 01613 02413 03213 10513 11313	}	—	ABCF, ABDG, ADEF	a, b, ac, ae	{ 84 78 71 63 64 55	— — — — ABCF ABCF, BCDE
10512 11312						
10512 11312	}	—	ABFG, ACDF	a, b, ac	{ 67 57	— —
Quarter-replicate designs: 64 treatment combinations						
00822 01622	}	—	ACF, AEH	a, e, ag	{ 24 18	— —

Table 5. Designs of type 2⁹

Table 5. Designs of type 2 ^a						
Scheme number	Generators of defining contrasts	Generators of interactions confounded with blocks	Generators for principal block	Multipliers for other blocks	D.F. for error	1st-order interactions confounded
Half-replicate designs: 256 treatment combinations						
00914 01714 02514 03314 04114	} <i>ABCDEFGHJ</i>	<i>ADH, AEJ, BDF, BEG</i>	<i>abde, achj, adjj, aegh</i>	<i>ab, ac, ad, cf</i>	195	—
188					—	
180					—	
171					—	
162					<i>CGH</i>	
04113	<i>ABCDEFGHJ</i>	<i>ADH, AEJ, BEG</i>	<i>abde, achj, adjj, aegh, cf</i>	<i>ab, ac, ad</i>	169	—
10614 11414 12214	} <i>ABCDEFGHJ</i>	<i>ADJ, BDF, CFH, ACFG</i>	<i>abdj, adfh, bcej, bdej</i>	<i>ab, ac, ad, ae</i>	170	—
160					<i>ACDE, ACDE, ACFG, DEFG</i>	
150						
11413 12213	} <i>ABCDEFGHJ</i>	<i>AEFG, CDEF, ABDEJ</i>	<i>abcf, abde, abgh, acdj, bj</i>	<i>ab, ac, ad</i>	167	—
155					—	
20312	<i>ABCDEFGHJ</i>	<i>ACDG, BCEH</i>	<i>ad, ag, be, bh, fj, abcf</i>	<i>ab, ac</i>	141	—
Quarter-replicate designs: 128 treatment combinations						
00923 01723 02523 03323	} <i>ADEFGH, BCDEGJ</i>	<i>ADJ, BHJ, CGJ</i>	<i>beh, cfj, achj, adef</i>	<i>af, ah, bc</i>	75	—
68					—	
60					—	
51					—	
10623 10622	<i>ABDEFH, BCFGHJ, ABDEFH, BCFGHJ</i>	<i>AEF, CEH, FJ, AEF, CEH</i>	<i>afj, bdg, cdh, aceg, afj, bdg, cdh, aceg, gj</i>	<i>ad, ae, gj, ad, ae</i>	51	<i>FJ</i>
54					—	

Table 6. *Designs of type 2¹⁰*

Table 6. <i>Designs of type 2¹⁰</i>						
Scheme number	Generators of defining contrasts	Generators of interactions confounded with blocks	Generators for principal block	Multipliers for other blocks	D.F. for error	1st-order interactions confounded
Quarter-replicate designs: 256 treatment combinations						
00(10)24 01824 02624 03424 04224	$\left\{ \begin{array}{l} ADEGHJ, \\ BCFGHK \end{array} \right.$	$\begin{array}{l} ACE, BCG, \\ ADK, AGH \end{array}$	$\begin{array}{l} bcej, bfg h, \\ dfjk, abchk \end{array}$	$\left\{ \begin{array}{l} ad, ae \\ aj, bf, \end{array} \right.$	$\left\{ \begin{array}{l} 185 \\ 177 \\ 168 \\ 158 \\ 148 \end{array} \right.$	$\begin{array}{c} — \\ — \\ — \\ — \\ AGH \end{array}$
04223 05023	$\left\{ \begin{array}{l} ABDEHJ, \\ BCFGJK \end{array} \right.$	$\begin{array}{l} CEFJ, CFHK, \\ ACHJ \end{array}$	$\begin{array}{l} bdg, fhj, acdf, \\ bcjk, cegh \end{array}$	$\left\{ \begin{array}{l} ad, ae, \\ ah \end{array} \right.$	$\left\{ \begin{array}{l} 155 \\ 143 \end{array} \right.$	$\begin{array}{c} — \\ — \end{array}$
10723 11523 12323	$\left\{ \begin{array}{l} ACDGHJ, \\ BCEFJK \end{array} \right.$	$\begin{array}{l} ADF, BFH, \\ BDK \end{array}$	$\begin{array}{l} cj, egj, adek, \\ aefh, bghk \end{array}$	$\left\{ \begin{array}{l} ad, ag, \\ ah \end{array} \right.$	$\left\{ \begin{array}{l} 164 \\ 152 \\ 139 \end{array} \right.$	$\begin{array}{c} — \\ — \\ — \end{array}$
Eighth-replicate designs: 128 treatment combinations						
00(10)33 01833 02633	$\left\{ \begin{array}{l} AEFGJ, \\ BDGHJ, \\ BCEFJK \end{array} \right.$	$\begin{array}{l} ACE, CDK, \\ EHK \end{array}$	$\begin{array}{l} bfg, acdj, \\ aefhj, acghk \end{array}$	$\left\{ \begin{array}{l} ae, ck, \\ dh \end{array} \right.$	$\left\{ \begin{array}{l} 65 \\ 57 \\ 49 \end{array} \right.$	$\begin{array}{c} — \\ — \\ CDK \end{array}$
10732	$\begin{array}{l} ABDEFH, \\ BCFGHJ, \\ ABCDFJK \end{array}$	$\begin{array}{l} ADF, BFJ \end{array}$	$\begin{array}{l} ad, bej, cgk, \\ egh, afjk \end{array}$	$\begin{array}{l} bf, cj \end{array}$	$\begin{array}{l} 40 \end{array}$	$\begin{array}{c} — \end{array}$

Table 7. *Designs of type 2¹¹*

Scheme number	Generators of defining contrasts	Generators of interactions confounded with blocks	Generators for principal block	Multipliers for other blocks	D.F. for error	1st-order interactions confounded
Eighth replicate designs: 256 treatment combinations						
00(11)34 01934	$\left\{ \begin{array}{l} ACDEL, \\ BCDHJ, \\ CFGJKL \end{array} \right.$	$\left\{ \begin{array}{l} ADH, BCE, \\ BDG, DEK \end{array} \right.$	$\left\{ \begin{array}{l} ahjl, acdeg, \\ bcfgl, cefjk \end{array} \right.$	$\left\{ \begin{array}{l} ae, bh, \\ fg, abd \end{array} \right.$	$\left\{ \begin{array}{l} 174 \\ 165 \end{array} \right.$	$\left\{ \begin{array}{l} - \\ - \end{array} \right.$
02734	$\left\{ \begin{array}{l} ACGJK, \\ EGHKL, \\ ABDEFGJ \end{array} \right.$	$\left\{ \begin{array}{l} FK, CEF, \\ DEG, EHJ \end{array} \right.$	$\left\{ \begin{array}{l} ahjl, abdgl, \\ acegj, defjk \end{array} \right.$	$\left\{ \begin{array}{l} aj, bd, \\ bf, acd \end{array} \right.$	156	FK
03534	$\left\{ \begin{array}{l} ABCFKL, \\ ACEHJK, \\ ADEFGJ \end{array} \right.$	$\left\{ \begin{array}{l} ACE, ADH, \\ AFJ, BFH \end{array} \right.$	$\left\{ \begin{array}{l} cegl, acfhk, \\ bdghk, abchjl \end{array} \right.$	$\left\{ \begin{array}{l} bl, ck, \\ dg, ej \end{array} \right.$	145	ABK
01933 02733 03533 04333	$\left\{ \begin{array}{l} AFHJK, \\ BDFHL, \\ CEGHKL \end{array} \right.$	$\left\{ \begin{array}{l} BEL, BFK, \\ CEJ \end{array} \right.$	$\left\{ \begin{array}{l} abkl, acegj, \\ adgh, \\ afhk, abcef \end{array} \right.$	$\left\{ \begin{array}{l} aj, bd, \\ ce \end{array} \right.$	$\left\{ \begin{array}{l} 173 \\ 163 \\ 152 \\ 140 \end{array} \right.$	$\left\{ \begin{array}{l} - \\ - \\ - \\ - \end{array} \right.$
10833 11633	$\left\{ \begin{array}{l} BDGKL, \\ CEHJL, \\ ACDFGJ \end{array} \right.$	$\left\{ \begin{array}{l} ADF, BEG, \\ CDL \end{array} \right.$	$\left\{ \begin{array}{l} af, bghj, dfhl, \\ egjk, chjkl \end{array} \right.$	$\left\{ \begin{array}{l} bk, cj, \\ dg \end{array} \right.$	$\left\{ \begin{array}{l} 149 \\ 136 \end{array} \right.$	$\left\{ \begin{array}{l} - \\ - \end{array} \right.$
Sixteenth replicate designs: 128 treatment combinations						
00(11)43 01943	$\left\{ \begin{array}{l} ACDEH, \\ BCDFJ, \\ ABCDGL, \\ BCEGJK \end{array} \right.$	$\left\{ \begin{array}{l} ACJ, AEF, \\ BCG \end{array} \right.$	$\left\{ \begin{array}{l} adehj, bceffj, \\ cghjk, defkl \end{array} \right.$	$\left\{ \begin{array}{l} acj, adf, \\ aeg \end{array} \right.$	$\left\{ \begin{array}{l} 54 \\ 45 \end{array} \right.$	$\left\{ \begin{array}{l} - \\ - \end{array} \right.$

Table 8. *Designs of type 2¹²*

Sixteenth replicate designs: 256 treatment combinations						
00(12)44 01(10)44 02844 03644 04444	$\left\{ \begin{array}{l} ACEHK, \\ BCFGJ, \\ DGJKL, \\ FHJLM \end{array} \right.$	$\left\{ \begin{array}{l} AEL, AGJ, \\ BHL, LM \end{array} \right.$	$\left\{ \begin{array}{l} acegk, adefj, \\ abcfjlm, \\ abcdefgh \end{array} \right.$	$\left\{ \begin{array}{l} ae, abc, \\ adk, \\ gjm \end{array} \right.$	$\left\{ \begin{array}{l} 163 \\ 153 \\ 142 \\ 130 \\ 118 \end{array} \right.$	$\left\{ \begin{array}{l} LM \\ LM, GHK \end{array} \right.$
00(12)43 01(10)43 02843 03643 04443	$\left\{ \begin{array}{l} ACEHK, \\ BCFGJ, \\ DGJKL, \\ FHJLM \end{array} \right.$	$\left\{ \begin{array}{l} AEL, AGJ, \\ BHL \end{array} \right.$	$\left\{ \begin{array}{l} gjm, bchm, \\ acegk, \\ adefj, abcfgl \end{array} \right.$	$\left\{ \begin{array}{l} ae, abc, \\ adk \end{array} \right.$	$\left\{ \begin{array}{l} 170 \\ 160 \\ 149 \\ 137 \\ 124 \end{array} \right.$	$\left\{ \begin{array}{l} - \\ - \\ - \\ - \\ - \end{array} \right.$
10944 11744	$\left\{ \begin{array}{l} ACDEFHK, \\ BCDEFGJ, \\ CDEJKM, \\ CDFGHL \end{array} \right.$	$\left\{ \begin{array}{l} ADG, CDF, \\ CGK, BJ \end{array} \right.$	$\left\{ \begin{array}{l} bjm, aeghk, \\ cfhkl, cdeghm \end{array} \right.$	$\left\{ \begin{array}{l} cd, cel, \\ acjl, ehjl \end{array} \right.$	$\left\{ \begin{array}{l} 129 \\ 116 \end{array} \right.$	$\left\{ \begin{array}{l} BJ, EG, FL, \\ ACH \\ BJ, EG, FL, \\ ACH, ADE \end{array} \right.$
10943 11743	$\left\{ \begin{array}{l} ACDEFHK, \\ BCDEFGJ, \\ CDEJKM, \\ CDFGHL \end{array} \right.$	$\left\{ \begin{array}{l} ADG, CDF, \\ CGK \end{array} \right.$	$\left\{ \begin{array}{l} bjm, ehjl, \\ aeghk, \\ cfhkl, cdeghm \end{array} \right.$	$\left\{ \begin{array}{l} cd, cel, \\ acjl \end{array} \right.$	$\left\{ \begin{array}{l} 133 \\ 119 \end{array} \right.$	$\left\{ \begin{array}{l} - \\ - \end{array} \right.$

Table 9. Designs of type 2¹³

Scheme number	Generators of defining contrasts	Generators of interactions confounded with blocks	Generators for principal block	Multipliers for other blocks	D.F. for error	1st-order interactions confounded
Thirty-second replicate designs: 256 treatment combinations						
00(13)54 01(11)54 02954	$\left\{ \begin{array}{l} ACEGK, \\ ADHJN, \\ ABCGJM, \\ ACDEFM, \\ ACHKLM \end{array} \right.$	$ACM, ADG, \\ AEL, AFN$	$acefgj, bdeghl, \\ cefhlmn, \\ dfghjkn$	$acn, bjn, \left\{ \begin{array}{l} dfn, \\ abdk \end{array} \right.$	$\left\{ \begin{array}{l} 149 \\ 138 \\ 126 \end{array} \right.$	— — —
03753 04553	$\left\{ \begin{array}{l} ACEGJ, \\ BDEKN, \\ BDFGM, \\ ABCFHK, \\ ABDHJL \end{array} \right.$	$ADF, BCDG, \\ BCEFH$	$adgk, bdfm, \\ ehjk, \\ cdfjn, cghlm$	$acl, cek, \left\{ \begin{array}{l} cfg \end{array} \right.$	$\left\{ \begin{array}{l} 121 \\ 108 \end{array} \right.$	GHK —
10(10)53	$ADHMN, \\ CEGKN, \\ BCDEFN, \\ BCGHJN, \\ BCKLMN$	$ADE, AFH, \\ BEF$	$bden, clmn, \\ gjkl, \\ defhmn, \\ abcd fkl$	$acn, adf, \\ ahj$	117	GJ

Table 10. Designs of type 2¹⁴

Sixty-fourth replicate designs: 256 treatment combinations						
00(14)63 01(12)63 02(10)63 03863 04663	$\left\{ \begin{array}{l} ACEHJ, \\ BCFGM, \\ ABDFHK, \\ FGHJKL, \\ JKLMNP, \\ ADEFGHN \end{array} \right.$	$ACF, ADH, \\ GP$	$achjm, dehl n, \\ glmnp, abdfjn, \\ cfjkl n$	$cfh, aefg, \left\{ \begin{array}{l} ahl p \end{array} \right.$	$\left\{ \begin{array}{l} 144 \\ 132 \\ 119 \\ 105 \\ 90 \end{array} \right.$	$\left. \begin{array}{l} \\ \\ \\ \\ \end{array} \right\} GP$
10(11)63	$ADEFG, \\ AGLMN, \\ BDEJP, \\ BEFHM, \\ CDFKP, \\ ABDEKL$	$ADP, BDK, \\ JM$	$aefgp, cfghn, \\ dfhkp, abcdhl, \\ egjlmn$	$cjp, hmn, \\ afkm$	99	JM

Confidence limits in the case of the geometric distribution

By KERMIT G. CLEMANS

U.S. Naval Ordnance Test Station, China Lake, California

INTRODUCTION

This paper presents central confidence belts for estimating m , the expected number of trials before failure, when X , the number of trials before failure, has a geometric distribution

$$\Pr(X = k) = p^k q \quad (k = 0, 1, 2, \dots),$$

where p and q are positive constants with $p + q = 1$. The expected value of X

$$E(X) = p/q = m$$

hence $p = m/(1+m)$ and $q = 1/(1+m)$.

This model is of interest in estimating the mean life of ordnance devices that have passed acceptance tests and have an operating life throughout which the probability of failure is constant. Consider a sample of N such devices randomly selected from a population and each subjected to a series of identical use cycles until it fails. Let X_i be the number of trials before failure of the i th device. If the X_i have identical geometric distributions, a sufficient estimator of the mean m is

$$\hat{m} = \sum_{i=1}^N X_i / N.$$

It is further known that $\sum_{i=1}^N X_i$ has a Pascal distribution (Feller, 1950, § 11.3), that is

$$\Pr\left(\sum_{i=1}^N X_i = k\right) = \binom{N+k-1}{k} p^k q^N \quad (k = 0, 1, 2, \dots),$$

where $p = m/(m+1)$ and $q = 1/(m+1)$. The distribution is bimodal with modes for $\sum X_i = m(N-1)-1$ and $m(N-1)$, $E(\sum X_i) = Np/q = Nm$, and $\text{var}(\sum X_i) = Np/q^2 = Nm(m+1)$. Then $E(\hat{m}) = m$, $\text{var}(\hat{m}) = m(m+1)/N$ and

$$\text{coefficient of variation of } \hat{m} = \sqrt{\left(\frac{m+1}{mN}\right)} \propto \frac{1}{\sqrt{N}}.$$

The cumulative Pascal distribution is

$$F(k) = \Pr\left(\sum_{i=1}^N X_i \leq k\right) = \sum_{j=0}^k \binom{N+j-1}{j} \left(\frac{m}{m+1}\right)^j \left(\frac{1}{m+1}\right)^N.$$

CONSTRUCTION AND USE OF CONFIDENCE BELTS

The belts of $1 - \alpha$ confidence for each N were determined by assigning m and then finding corresponding values k_1 and k_2 where

$$F(k_1) \leq \frac{1}{2}\alpha < F(k_1 + 1)$$

and

$$F(k_2) < 1 - \frac{1}{2}\alpha \leq F(k_2 + 1).$$

For the purpose of constructing curves a k' for $\frac{1}{2}\alpha$ and a k'' for $1 - \frac{1}{2}\alpha$ were calculated by linear interpolation between k_i and $k_i + 1$ and the points $(k'/N, m)$ and $((k'' + 1)/N, m)$ were plotted.

The summation was done by the IBM 704 Computer and the program will be kept available. Belts of different confidence for the sample sizes used in this article are available from the author. Charts I and II give two-sided 90 and 98 % central confidence limits, respectively, for m , the mean number of trials before failure, as a function of k/N , the observed mean number of trials before failure in samples of size N . The charts also give one-sided 95 and 99 % confidence limits.

In general, these confidence belts can be used in the same way as the binomial confidence belts of Clopper & Pearson (1934) or Crow's tables (1956), to estimate confidence limits or to plan in advance the size of sample necessary for a given degree of accuracy.

The confidence intervals estimated from Charts I and II are in agreement with those computed by Anscombe's method if his extra correction terms are used when necessary (Anscombe, 1948).

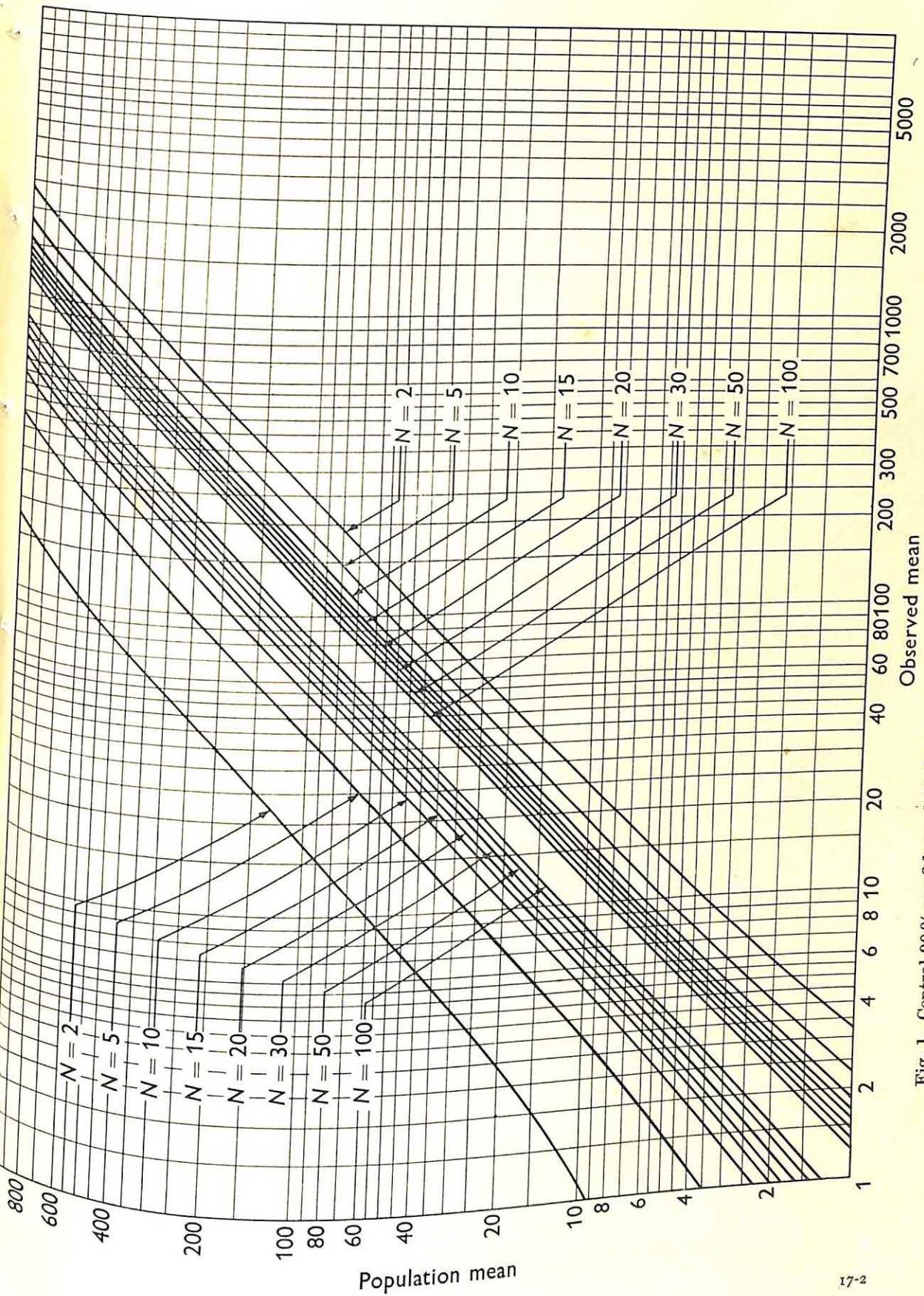


Fig. 1. Central 90 % confidence belt for the mean of a geometric distribution based on a sample of size N .

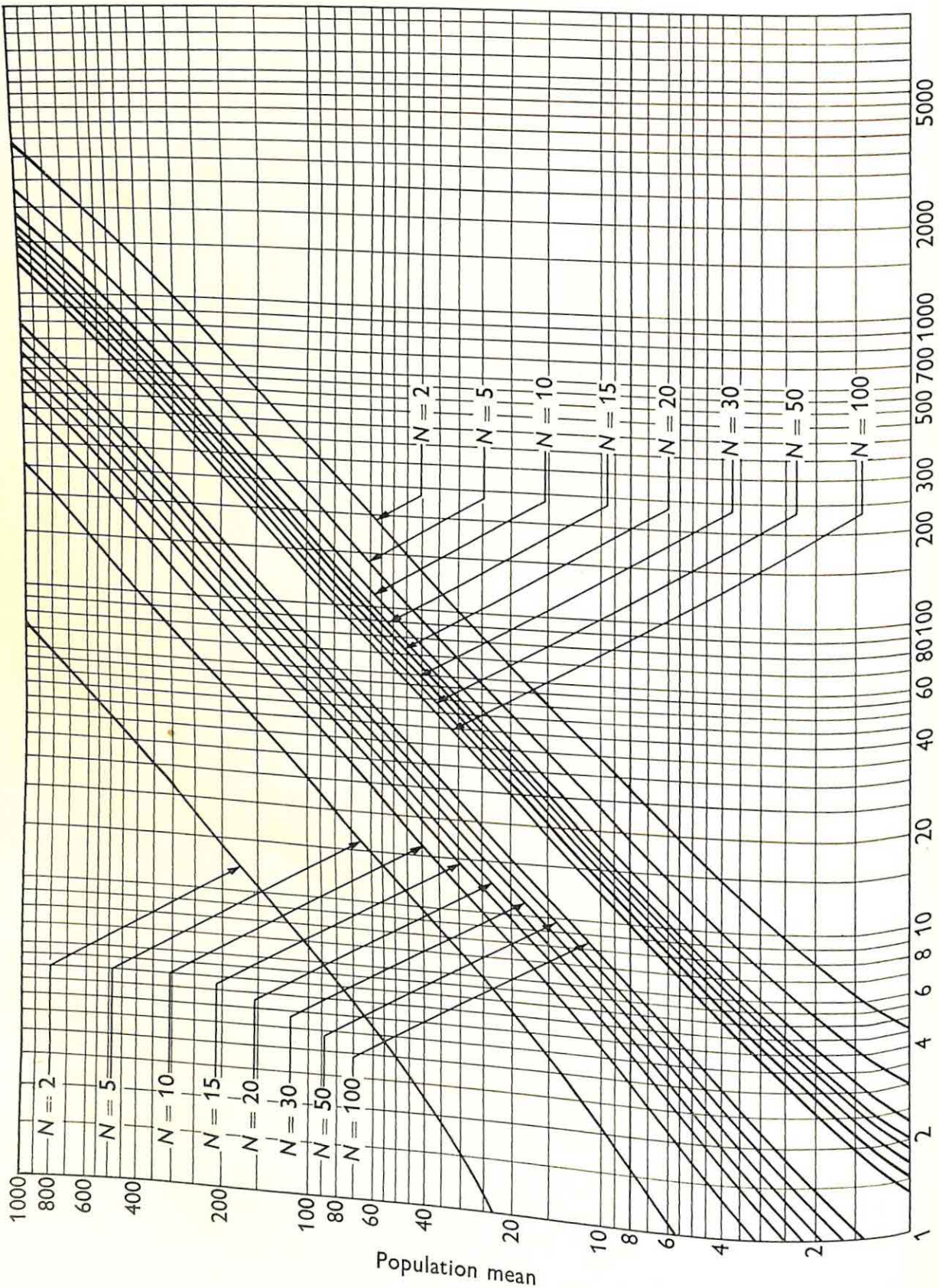


Fig. 2. Central 98.0% confidence belt for the mean of a geometric distribution based on a sample of size N .

Example 1. What are the 90 % confidence limits for m , the expected number of trials before failure, if $N = 30$ units operated for a total of $k = 15,000$ trials? The desired confidence limits are the ordinates of the intersection of the two curves for $N = 30$ in Chart I and the abscissa $k/N = 500$, i.e. 385-700.

Example 2. What value will the mean exceed with 95 % confidence in Example 1? The desired value is the lower limit of the central 90 % confidence band, that is 385.

INTERPOLATION AND EXTRAPOLATION IN THE CHARTS

Interpolation with respect to sample size may be carried out linearly on the logarithms of the population mean with arguments N^{-1} .

Extrapolated confidence limits for the population mean for sample size $N > 100$ are found from the equation

$$\log m = \log(k/N) + 10N^{-1} |\log m_{100} - \log(k/N)|,$$

where m_{100} is the corresponding (upper or lower) limit for sample size 100 at abscissa k/N .

Extrapolated population mean confidence belts for large sample means (k/N) are

$$m_U(N) = (1000/a)(k/N)$$

$$m_L(N) = (1000/b)(k/N),$$

and

where a and b are the abscissas at population mean 1000 for the corresponding sample size N ; see Tables 1 and 2. To obtain values of a and b for values of N not tabled, interpolate linearly on values of $\log a$ and $\log b$, using N^{-1} as argument.

Table 1. Value from 90 % confidence belt

N	2	3	4	5	10	15	20	30	50	100	∞
a	177.0	272.0	341.1	393.6	542.3	616.2	662.6	719.7	779.2	841.4	1000
b	2373	2100	1940	1832	1572	1460	1395	1320	1245	1173	1000
c	5.23	3.20	2.42	1.99	1.21	0.938	0.793	0.632	0.480	0.336	0

Table 2. Values from 98 % confidence belt

N	2	3	4	5	10	15	20	30	50	100	∞
a	73.57	144.8	205.3	255.4	412.7	498.2	553.9	624.6	700.5	782.1	1000
b	3324	2807	2516	2326	1883	1702	1598	1480	1366	1259	1000
c	13.29	6.55	4.47	3.49	1.89	1.42	1.18	0.925	0.696	0.484	0

Example 3. Find the size of sample necessary to ensure that the 90 % confidence interval for the population mean is not longer than 0.50 of the sample mean. This means the magnitude of the distance across the 90 % confidence belt measured along an abscissa should be $c = 0.500$ of that abscissa. Find c in Table 1, where

$$c = 1000 \left(\frac{1}{a} - \frac{1}{b} \right),$$

and a and b are the abscissas at $m = 1000$ of the confidence belt.

The closest value is $c = 0.480$ for $N = 50$. A more accurate estimate of the sample size required may be found by interpolating between values of N^{-1} , using as arguments logarithms of c .

$$\frac{\log(0.500) - \log(0.480)}{\log(0.632) - \log(0.480)} = 0.148390,$$

$$30^{-1} - 50^{-1} = 0.041153,$$

$$N^{-1} = 50^{-1} + (0.041153)(0.148390) = 0.147528.$$

This value of N^{-1} is between 45^{-1} and 46^{-1} ; hence the sample size is taken as 46.

REFERENCES

- ANScombe, F. J. (1948). The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, **35**, 246-54.
- CLOPPER, C. J. & PEARSON, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, **26**, 404-13.
- CROW, E. L. (1956). Confidence intervals for a proportion. *Biometrika*, **43**, 423-35.
- FELLER, W. (1950). *An Introduction to Probability Theory and Its Applications*, **1**. New York: John Wiley and Sons, Inc.

Note regarding Statistical Research Monographs

The Institute of Mathematical Statistics and the University of Chicago have established a series of publications entitled *Statistical Research Monographs*.

The primary purpose of this series is to provide a medium of publication for material of interest to statisticians that is not ordinarily provided for by existing media. It will help fill the gap between journal articles and textbooks or treatises. Among the kinds of publications envisaged are:

New research results too lengthy for the usual journal article. In particular, authors will have ample scope for detailed exposition of their findings.

Research results of interest in both theoretical and applied statistics. At present authors of such material frequently find it necessary to publish part of their results in a theoretical journal and part in an applied journal.

Expository monographs in particular areas of statistics.

Discussions of statistical problems and techniques in particular areas of application.

Every attempt will be made to maintain the highest standards of scholarship.

The Editorial Board consists of David Blackwell (University of California), William G. Cochran (Harvard University), Henry E. Daniels (University of Birmingham), Leo A. Goodman (University of Chicago), Wassily Hoeffding (University of North Carolina), Jack C. Kiefer (Cornell University) and William H. Kruskal (University of Chicago).

Authors are invited to send manuscripts and correspondence concerning the series to Leo A. Goodman, Department of Statistics, University of Chicago, Chicago 37, Illinois.

REVIEWS

The Logic of Scientific Discovery. By KARL R. POPPER. London: Hutchinson and Co. (Publishers) Ltd. 1959. Pp. 480. 50s.

Prof. Popper's *Logik der Forschung* was first published in 1934. It was recognized from the first as a most important book and has had a profound effect on philosophers, though not very much on statisticians, except perhaps those who have been lucky enough to be able to attend Popper's courses on Scientific Method at the London School of Economics. Rather surprisingly, it has not previously appeared in an English translation. This new version is especially welcome at a time when scientific discovery is proceeding so rapidly and, in particular, when statisticians are still trying to put their inferential procedures on a proper formal basis.

Popper has been working on his problems since 1934, and although his views are not modified in any fundamental respect he now has much more to say. One imagines that he sat down to this translation with the idea of adding his maturer views in a few footnotes and appendices. But the new material nearly got out of control. Of the 480 pages of this book 156 are devoted to new appendices, twelve in number; and his work on what he calls the propensity interpretation of probability has been hived off into a new book (*Postscript: After Twenty Years*) which is still to appear.

To try to present the essentials of this book in a short review is perhaps ambitious; but I make the attempt in the hope that readers of this journal may be encouraged to tackle the original work.

Popper rejects the Baconian idea of induction in science, under which the study of a mass of factual observations led to the expression of a 'vintage' of universal law. He draws a sharp distinction between the process of conceiving a new idea and the examination of it. How the mind thinks of hypotheses is not his primary concern, and in any case is not a logical process in the strict meaning of the word 'logic'. He is, in this book, mainly interested in the deductive testing of a theory or a hypothesis. Further, he admits a system as scientific only if it is capable of being tested by experience. Thus the principal criterion whether a system is scientific or not is that it is testable or falsifiable.

It is not always possible to compare degrees of falsifiability, but when it is, Popper says that, of two statements, the one which is less falsifiable is the more probable. But this is a kind of probability, 'absolute logical probability', or *a priori* probability which will be new to the statistician and is not to be confused with numerical chance. The more a statement forbids, the more it says about the world of experience, and hence the greater its empirical content, its falsifiability and hence its testability; but the smaller its absolute logical probability because, to put it shortly, it forbids a larger class of possible events.

Popper then equates the *degree of falsifiability* to the *simplicity* of a hypothesis, and has some very interesting comments on the concept of simplicity itself. Thus he arrives at the schema

testability = high prior improbability = paucity of parameters = simplicity.

This is to be compared with Jeffreys and Wrinch, who effectively would have

high prior probability = paucity of parameters = simplicity.

At this point, as he himself recognizes, Popper seems to be heading for trouble. For if the parameters in a hypothesis may take an infinite set of values the prior probability of any particular hypothesis is zero; and, in fact, it was to avoid this difficulty that Jeffreys set up his method of approach to prior probability. But Popper has an answer to this. The assertion that a hypothesis has a non-zero probability is not itself testable and, whatever its metaphysical value, has no value or validity in science. We should therefore not be concerned with the probability of hypotheses but with their *degree of corroboration*. It is possible to go on learning about a theory, to test it more and more, and to increase its degree of corroboration, without increasing its logical probability, which remains at zero.

Popper then has to set up a theory of corroboration (Bewährung). In earlier work he and, following him, Carnap have used the words *degree of confirmation*, but this expression he now discards as misleading (and, in fact, as having misled some of his critics, including Carnap). Popper distinguishes between 'probability' in the sense of relative frequency, 'logical probability' in relation to statement and 'corroboration'. In statistics he has accepted only the relative frequency interpretation of probability, though he has since modified this viewpoint by formulating his propensity theory. Within the framework of frequency theory he has in the first instance to dispose of two frequential problems: (a) the problem of the random sequence; this he does, as I understand it, by modifying the field of

possible tests for randomness by making it denumerable; (b) the problem that no assertion about a probability is testable because (again putting it roughly) anything can happen by chance; this he overcomes by an assumption rather akin to d'Alembert's, that events of small probability simply do not yield occurrences in Nature.

These difficulties behind him, he can proceed to a theory of corroboration. It is not so much the number of corroborating instances which determines the degree of corroboration as the severity of the various tests. The greater the possible severity, the greater the falsifiability; and hence the simpler hypothesis is more corroborable. The degree of corroboration of x by y is then not the same thing as the probability of x given y ; and in fact it is possible to construct examples in which x_1 is strongly supported by y and x_2 is strongly undermined by it, but for which $P(x_1|y) < P(x_2|y)$.

Nevertheless, the degree of corroboration is defined as a function of probabilities. Popper's original method of doing so may, perhaps, require modification, but at least he showed that a consistent system obeying his axioms is possible. He thus measures the rationality of accepting the guess which generated the hypothesis, and this seems to be about as far as numerical evaluation of the acceptability of a theory can go.

The book contains a good deal of technical matter which will be mainly of interest to the logician, and some replies to criticism which will interest only those who are aware of the work of the original critics. Moreover, a final judgement of the system must await the sequel volume. Nonetheless, this book remains an outstanding contribution to the problem of the utilization of probabilistic ideas in the evaluation of hypotheses and will repay detailed study.

M. G. KENDALL

Experimental Designs in Industry. [Symposium held November 1956 at North Carolina State College.] Edited by V. CLEW. New York: John Wiley and Sons Inc. London: Chapman and Hall Ltd. 1958. Pp. xi+268. 48s.

This book contains nine papers on experimental design read at a symposium held in November 1956 at North Carolina State College. The papers are grouped in two Parts. Part I contains four substantial papers (192 pages in all) containing, between them, most of the experimental designs which are known to have been found useful in industrial research. Part II contains five rather briefer papers on applications of experimental design techniques to specific industrial problems. The book concludes with an extensive twelve-page bibliography. There are also copious references given with some individual papers.

The outstanding feature of the first part of the book is the coherence and continuity of the four papers contained in it. This is an unusual quality in a symposium, possibly due to luck, but more probably reflecting some care in planning. In the first paper (Basic Experimental Design), the editor gives a clear, competent introduction to the general theory and application of all the simpler designs, including an account of the general linear hypothesis approach. This is followed by papers on: complete factorials, fractional factorials, and confounding (R. L. Anderson); simple and multiple regression analysis (R. J. Hader & A. H. E. Grandage) and experimental design for exploring response surfaces (G. E. P. Box & J. S. Hunter). These develop, in a natural way, the basic structure described in the first paper.

It would be quite possible for the unaided, mathematically inclined, student to obtain a good idea of modern experimental design from these four papers. They do not, of course, form an 'ideal' text-book—in particular, the theoretical bases underlying the numerical procedures are not always sufficiently clearly described for pedagogical purposes—but they do describe, in a clear unambiguous fashion, the procedures which are actually used.

The papers on specific industrial applications naturally form a rather less closely knit group. They are, however, of considerable intrinsic interest, enhanced by the fact that they give actual designs, data and analyses. The topics covered include examples of incomplete block design, fractional replication, and response surface design; and a short paper on 'Design in ordnance experimentation'.

This is a book which should prove of real value to serious students of statistics and others wishing to learn something of the techniques and outlook of modern experimental design. It is an unpretentious book (typescript with a number of surprising slips in spelling) and does not claim more for its subject than is justified. It is also quite inexpensive, by present-day standards.

N. L. JOHNSON

Analysis of Multiple Time-Series. [No. 1 of Griffin's Statistical Monographs and Courses.] By M. H. QUENOUILLE. London: Charles Griffin and Co. Ltd. 1957. Pp. 105. 24s.

This is the first in a series of monographs on specialized topics which are being edited by Prof. M. G. Kendall (the second was reviewed in the last issue of this journal).

Mr Quenouille's book is remarkable in that much of what it contains is new work developed over a relatively short period. The author is concerned with an extension to stationary multiple series of the correlation properties, which are well known in the case of single series and also makes a start with the associated inference problems, confining himself almost entirely to large sample theory. Specifically, the models which are considered are linear of the form $\mathbf{H}(D)\mathbf{X}_t = \mathbf{G}(D)\epsilon_t$, where $\mathbf{H}(D)$ and $\mathbf{G}(D)$ are matrices whose elements are algebraic functions of the shift operator D , $\mathbf{X}_t$ is a vector of p elements representing the simultaneous values of p series at time t and ϵ_t is a $p \times \infty$ matrix of uncorrelated variables extending backwards in time. One would have liked to see at the outset a statement concerning the uniqueness of this representation in the spirit of Wold's theorem for single series.

In the case where $\mathbf{G}(D) = \mathbf{I}$, the unit matrix, the correlation properties of the $\mathbf{X}_t$ are determined by the roots of the determinantal equation $|\mathbf{H}(D)| = 0$ and these may be real or imaginary in pairs. It is shown that much simplification is obtained if a transformation is made to a canonical form in which $\mathbf{H}(D)$ is diagonal with elements which are the factors of $|\mathbf{H}(D)| = 0$. In the case where these factors are real, the transformed model is that of a multivariate Markoff scheme viz. $\mathbf{A}_0\mathbf{X}_t^* + \mathbf{A}_1\mathbf{X}_{t-1}^* = \epsilon_t^*$ and these are given special treatment in what follows.

Chapter 3 is concerned with the problem of Identification, namely, that the parameters of the scheme cannot be uniquely determined from the covariance structure. This is true even for the model cited earlier, but the situation is made more complex by the effect of superposed error. In the case of a single series this is equivalent to mixing the series with an independent random series of unknown variance. Clearly, one could make the situation more hopeless by adding an autocorrelated series with unknown autocorrelation structure, in which case it would follow that the two series could not be separated without further knowledge or specification.

Chapter 4 deals with inference problems and the methods which are developed are illustrated by means of artificially generated multiple series. Chapter 5 is devoted to miscellaneous topics and is intended to demonstrate how the analysis may be complicated in practice due to the omission of relevant variables, moving average residuals (i.e. $\mathbf{G}(D) \neq \mathbf{I}$), superposed error, non-linearities, trends and short series of observations. A very interesting method of trend analysis is proposed in which the differences of the correlations rather than those of the original series are used. Next follows chapters on estimation and goodness-of-fit tests and finally a last chapter in which a practical example (five American economic series) is subjected to the various techniques developed earlier.

It is unfair perhaps to criticize a book which explores so much new ground, but it is necessary to draw attention to the fact that it should not be read uncritically. Thus, the use of criteria of the form $\mathbf{C}_s\mathbf{C}_{s-1}$, where $\mathbf{C}_s$ is the covariance matrix corresponding to lag s , as a means of deciding on the order of the scheme is clearly unsatisfactory for moving average residuals. In the case of a single series these matrix quotients are the partial serial correlations and these are zero beyond the k th for a k th order autoregressive scheme but never zero if the residuals are moving averages. The chapter on estimation contains several surprising statements, for example, that the lower serial correlations form a sufficient set for the estimation of the parameters of a moving average scheme. Again, it is surprising that in effect of mean corrections on the serial correlations is given such little space, despite the fact that in small samples they may have considerable effect on significance tests and estimation. The relations between continuous and discrete schemes given on pp. 48-50 are interesting, but not very useful since one is rarely in a position to say that one is sampling from a 'continuous' scheme governed by a specified stochastic differential equation—in spectral analysis it is necessary to ensure that there are no oscillations between the sampling points, using some preliminary filtering where necessary.

A general point which may occur to other readers is that the approach to time-series which is pursued in this monograph is entirely empirical—this is no doubt related to the author's interest in particular subject-matter fields where no other approach seems possible. In contrast to this it is possible in electrical engineering applications to specify the $\mathbf{X}_t$ and ϵ_t in such a way that the problem of identification disappears. Also, a linear model is used only when it is physically meaningful or else as an exploratory approach before more realistic non-linear models are used. It is not difficult to predict that it will become a standard book worthy of serious study for years to come.

G. M. JENKINS

Fundamentals of Statistical Reasoning. [No. 3 of Griffin's Statistical Monographs and Courses.] By M. H. QUENOUILLE. London: Charles Griffin and Co. Ltd. 1958. Pp. 169. 24s.

'The practical man seeking a step-by-step explanation of how to carry out statistical calculations will not find it here. Nor will the mathematician find proofs of the formulae involved in deriving such procedures. If either should, however, be interested in the foundations of statistics, in the principles rather than the proofs or practices, this monograph is aimed at him.' (From the author's preface.)

After this introduction, the present reviewer was somewhat disappointed to find that the first two parts (Basic Concepts, Deductive Methods) have been devoted to an exposition of the *mathematical* concepts used in statistical theory, such as is covered by the first part of Anderson and Bancroft's *Statistical Theory in Research*. There is very little discussion of the relation of these concepts to statistical practice in the real world. The treatment is superficial and in parts inaccurate. One example must suffice: thus 'independence' (of events) is never defined, and the word is first used in the phrase 'independent exhaustive possible events of a single trial'.

A great improvement is apparent in the final part (Inductive Principles) which occupies over half the book. The Neyman-Pearson theory is sketched, and the work of Jeffreys and of Wald is mentioned, but the main emphasis is on an outline of a development of fiducial theory from axioms appropriate to a measure of belief. Some rules are given as to when joint fiducial distributions can be found and integrated; these appear to give a self-consistent theory, but there is still some difficulty in deciding whether the rules are satisfied in any given case. It is to be hoped that Mr Quenouille will continue to develop his line of approach.

Professional statisticians (who should be both mathematicians and practical men) will be stimulated by this final part, and plain mathematicians may be intrigued; but Mr Quenouille's practical men are likely still to require convincing that the theory outlined has any relevance to their problems.

C. L. MALLOWS

Multivariate Correlational Analysis. By P. H. DuBois. New York: Harper and Brothers. 1957. Pp. 202. \$4.50.

Statisticians of the sophisticated sort are nowadays inclined to take a condescending attitude to correlation analysis and to attempt instead to operate as far as possible within the framework of regression and variance analysis. Indeed, it is hard to imagine what would have become of some of them if Fisher had never existed. Not so the psychologists, however, many of whom appear to be still contentedly following in the wake of Galton, Pearson and Yule.

The prime purpose of the present book, written for non-mathematicians, is to set out systematic methods of calculating Yule-type partial and multiple correlations. Its subsidiary purpose is to give a brief account of, and to relate to correlation analysis, the methods of factor analysis developed by Spearman and Thurstone. It is regrettable that scarcely any mention is made of orthodox modern statistics, and the distinction between population and sample is almost completely ignored. Less than a page is devoted to sampling distributions and hypothesis testing. Consequently, although the book will no doubt be of value to psychological research workers it is not likely to be very helpful to statisticians in other fields.

The book's chief merit is its demonstration of compact computing methods similar to those of Crout and Dwyer for the systematic calculation of correlation and regression coefficients by pivotal reduction of the correlation matrix. This is done with judicious mixing of partial correlation and multiple regression theory in a manner much simpler than straightforward application of Yule's theory. The author shows how partial variances and covariances may be identified at each stage of the reduction and shows how these may be combined to give correlation and regression coefficients of various kinds.

An important but irritating problem in this field is the choice of a 'best' subset of independent variables in a regression model, 'best' in the sense of maximizing the multiple correlation for subsets of a given size. For example, in psychological work we might have the results of twenty or more tests at our disposal from which we wish to select the five or six which best predict the dependent variable. It seems amazing that no one has produced an algorithm which under suitable conditions will automatically yield the best subset, apart from routine fitting of all possible regressions which will normally be out of the question because of the enormous number of possibilities involved. The customary expedient is to fit independent variables successively, taking first the one which is most highly correlated with the dependent variable, then that which in combination with the first gives the highest

regression sum of squares, and so on. This method is recommended in the book under review and a systematic computing procedure is given. As the author points out, however, not only is there no guarantee that the method will yield the best subset, it is easy to construct cases where it definitely does not. A defect of the presentation is that no rule is given for advising when to stop adding further independent variables. In this respect the presentation may be compared unfavourably with that of Wherry and that of Summerfield and Lubin (*Psychometrika*, 1951). Essentially, Wherry recommends stopping when the contribution to the regression sum of squares, due to including a further variable, is less than the estimated residual variance, while Summerfield and Lubin recommend stopping when the increment to the regression sum of squares is non-significant by the F -test. As an additional safeguard Summerfield and Lubin suggest checking that the difference between the regression sum of squares given by the reduced set of variables and the regression sum of squares given by the complete set of independent variables is small enough to be attributable to chance. These suggestions have obvious merits and it would have been helpful to have had them discussed.

Although the author's account of factor analysis is useful in that it ties it up with the earlier treatment of correlation and regression analysis, it seems altogether too short to give one a clear insight into this mysterious subject.

J. DURBIN

Introduction to Statistical Inference. By J. C. R. LI. Michigan: Edward Bros. Inc. 1957. Pp. 553. \$7.50.

The publishers state that this is a non-mathematical exposition of the theory of statistics written for experimental scientists. The mathematics are in fact kept to a minimum—there is just enough to enable necessary formulae to be written down succinctly and no attempt is made at proofs. The necessary theorems are stated as articles of dogma and sampling experiments are used to illustrate their validity. In this fashion Prof. Li covers the distributions of the sample mean and variance, t , F , the variance ratio distribution, the one-way classification in the analysis of variance, randomized blocks, linear regression, analysis of covariance, factorial experiments, transformations and distribution free methods.

The level of exposition is remarkably consistent and close to the mathematical minimum aimed at throughout. The writing is clear and the ideas are set out with simplicity. The reviewer feels that the paucity of the data of experience may be a handicap to the experimental scientists, since they tend to lean heavily on actual data rather than fictitious. However this is possibly a matter of opinion, and within its limits the book can be recommended.

F. N. DAVID

Teach yourself Statistics. By R. GOODMAN. London: English Universities Press Ltd. 1957. Pp. 238. 7s. 6d.

As its title implies this book is an introduction to statistics for readers with a minimum of mathematics. (Elementary calculus and determinants are the most advanced techniques used and when such techniques are used reference to appropriate texts is often given.) There is a wealth of worked examples, but the casual reader will need a desk calculating machine by him if he is to run through them himself. The field covered is the usual basic minimum: presentation of data, measures of frequency distributions, elementary probabilities, binomial, Poisson and univariate and bivariate normal distributions, regression and correlation, rank correlation, least squares, partial correlation, sampling distributions of means, etc., elementary estimation, the t - and F -distributions, elementary testing, analysis of variance, testing in regression analysis, χ^2 -tests of goodness-of-fit and contingency tables.

Whilst the topics and the order of presentation are much as in older elementary texts (Yule and Kendall springs to mind) the mode of presentation is not, at least in the earlier chapters. The author being faced with the fundamental problem in such a book, of relating the theory of probability in an elementary text of the early nineteenth century. For instance, he defines probability (pp. 34–35):

(1) If the situation is such that we can distinguish n "equally probable" outcomes of a set of circumstances C and if these are all the possible outcomes of C then if m of these outcomes entail the occurrence of an event E we say that the appropriate estimate of the probability of E given C , $p(E; C)$ is given by

$$p(E; C) = m/n.$$

(2) Otherwise, if in n occurrences of a set of circumstances C the event E has occurred m times the appropriate estimate of the probability of E given C is

$$p(E; C) = m/n.'$$

And again, if in N trials, the score x_i is obtained n_i times, $i = 1 \dots n$ he states (on p. 43)

$$E(x) = \lim_{N \rightarrow \infty} \left[\sum_{i=1}^n n_i x_i / N \right],$$

(allowing for what must be a misprint).

Whilst such statements as these are to some extent mitigated by introductory discussion, it must be said that the arguments used are often not only loose but such as seriously and unnecessarily to confuse the student.

There is another class of loose argument also used which is due to the low level of mathematics assumed in the reader, namely 'suggestive' or heuristic mathematical proofs and these are relatively venial and in any case unavoidable. Indeed, they contribute to the reader's understanding if he is under no illusions as to their nature. In this book, however, they are often so indirect, intricate and tedious that they are of doubtful utility. Further, they use such principles as, for instance, the uniqueness theorem for moments, without even an indication that this requires proof, and so beg the question in a hidden fashion. This sort of treatment can only unnerve the intelligent reader who has a limited knowledge of mathematics and lead him to mistrust his own judgement.

The book lacks an index, which is a serious shortcoming when important principles and techniques are introduced in an *ad hoc* manner, as and when they are needed for a particular application (as for instance, criteria such as bias, efficiency, etc. of estimators or the principles of testing statistical hypotheses).

The concepts of sufficiency, maximum likelihood estimation or power of a test do not seem to be mentioned.

D. E. BARTON

Concise Tables for Statisticians. By K. C. S. PILLAI. Manila: The Statistical Center, University of the Philippines. 1957. Pp. 50. \$1.50 plus postage.

This is a book of eighteen short tables on 33 octavo pages preceded by 14 pages of introductory matter. The tables fall into three groups. The first consists of seven basic tables: the normal density function and its integral, the percentage points (95 and 99 %) of the χ^2 , t , F and r (the sample correlation for $\rho = 0$) and a table of random numbers. The next group consists of eight tables of analogous functions based on normal range (rather than sample s.d.). The last group contains two tables of percentage points of the largest root (table 16) and the sum of roots (table 17) of the matrix generalization of the variance ratio for multivariate analysis. In the language of Foster and Rees, table 16 consists of the percentage points (95 and 99 %) of the generalized β distribution $I_x(k; p, q)$ for $k = 2, 3, 4, 5$ at a very much coarser set of intervals than Foster's and Foster & Rees's tables. Table 17 is an approximation based on fitting Pearson curves.

The introduction, in so far as the first two groups of curves are concerned, consists of little more than specification of the function tabled, but the multivariate tables are more fully introduced with worked examples of application. No indication is directly given of the method of computation used (except for table 17 as noted) nor of what methods of interpolation are appropriate. References are given, however, to the publications in which each table (or part of it) appeared.

D. E. BARTON

Functions of Real and Complex Variables. By W. F. OSGOOD. New York: Chelsea Publishing Company. 1958. Pp. 407 + 262. \$4.95.

This book—a reprint, in one volume, of two books published in 1936—deals with those branches of the theory of functions which are normally taught in a B.Sc. Special course in mathematics, and with some others, such as Fourier series, Fourier integrals, existence of solutions of differential equations, analytic continuation, Picard's theorem, and the more advanced theory of conformal mapping. The quality of the book is, in my opinion, fairly indicated by that of the following excerpt, in which the terms 'cluster point' and 'point of condensation' mean 'limit point', and 'upper limit' means 'exact upper bound':

THE WEIERSTRASS-BOLZANO THEOREM. *An infinite linear point set which lies in a finite interval, has at least one cluster point.*

Consider the points x of the line, such that, if x' be any one of them, only a finite number of points of the given set (in particular, none whatever) lie to the left of x' . This point set is bounded from above. It has, therefore, an upper limit, G . This point, G , is a point of condensation. For otherwise it would be an isolated point of the set, or not belong to the set at all. In either case, there would be a point x to the right of G .

T. ESTERMANN

An Introduction to Combinatorial Analysis. By JOHN RIORDAN. New York: John Wiley and Sons Inc. London: Chapman and Hall Ltd. 1958. Pp. x + 242. 68s.

The great name in combinatorial analysis is undoubtedly that of the late Major Macmahon. In two books of unrivalled ingenuity and combinatory insight he delineated the theory as we know it today, knitting together the early work of De Moivre, the generalizations of Laplace and much of the elementary work in the field of number theory. The difficulty, and it is a real one, in reading Macmahon, is the very generality of the theory because although it is all there it is not easily apparent how any particular piece of theory may be applied in any given circumstance.

The work of R. A. Fisher in the combinatorial field, both in sampling moments and in experimental design, has made it essential that the worker who would research in these fields should have some knowledge of the combinatorial approach. Previously there was no alternative but to study Macmahon: Prof. Riordan has now put us all in his debt by writing a treatise which will make the approach to Macmahon less painful and to a certain extent will supplement his work.

The topics covered are permutations and combinations, generating functions, De Moivre's theorem sometimes attributed to Boole and often to Whitworth, cycles of permutations, occupancy, partitions, compositions, trees and networks, and restricted permutations. There is no tie-up with statistical methods or with probability theory. Prof. Riordan's book should, however, be made required reading for all who attempt to work in the field of so-called non-parametric methods. The standard of exposition is high and it is very easy to learn from. In the opinion of the reviewer this is an excellent treatise which fills a gap in the mathematical armour of the statisticians.

F. N. DAVID

Calculus of Finite Differences. By GEORGE BOOLE. [Edited by J. F. MOULTON.] New York: Chelsea Publishing Co. 1958. Pp. viii + 336. \$4.95.

George Boole, in 1860, when he was professor of mathematics at Queen's College, Cork, wrote an elementary treatise on finite differences. After his death his book was revised by J. Moulton of Christ's College, Cambridge, in 1872. There is no record of the date of the third edition but the Chelsea Publishing Company note that the present (fourth) edition differs from the third in notational changes only.

This book, refreshingly free from the modern trend towards unnecessary mathematical rigour, has stood the test of time well. Generations of English mathematicians, statisticians and actuaries, have been reared on it, and that it has now crossed the Atlantic can only be a source of satisfaction. For those who are not familiar with 'Boole' the topics covered in the first part are Interpolation, Quadrature, Summation of Series, Bernoulli numbers, Convergency, Boole's limit for the remainder of a series. The second part is devoted mainly to difference equations. There are many examples in each chapter and a set of answers is given at the end of the book. To re-read this book after a number of years is to become aware of how little advance has been made in manipulative mathematics.

'Boole' can be recommended to students of all types and ages. The mathematical approach is simple and the amount of prerequisite learning is negligible. At the price, it is—by modern standards—in the bargain field.

F. N. DAVID

A Treatise on the Analytic Geometry of Three Dimensions. (Seventh edition.) Vol. I. By GEORGE SALMON [Edited by C. H. ROWE.] New York: Chelsea Publishing Co. 1958. Pp. xxiv + 470. \$4.95.

The introduction of the geometrical method in statistics by R. A. Fisher has been followed by many derivations and proofs in distribution theory based on this type of approach. The generalization from three to n dimensions is not always skilfully or accurately done, but the fact that it is done at all makes it a necessity for the modern statistics student to have at least a nodding acquaintance with three-dimensional geometry. There are several books in current use and in the reviewer's opinion there is no reason why this, the seventh edition of Salmon's book, should be preferred to them. This being said, it must also be remarked that this is an excellent book.

The date of the original edition of this work is not stated, but it was revised in 1911 for the fifth edition and again in 1927 for the seventh edition here reprinted. Salmon, who appears to have been an inveterate writer of text-books, almost rivalling Todhunter, wrote his treatise on conic sections in the 1870's so presumably this present work will date from about that time. The topics covered

are co-ordinates and their transformation, geometrical interpretation of equations, quadriplanar co-ordinates, properties of quartics and their classification, reciprocation and duality, foci and confocal surfaces, cones, and the general theory of surfaces. One or two examples and exercises are inset at intervals in the text. The index, contrary to most text-books, is both usable and informative.

F. N. DAVID

Probleme der Statistischen Methodenlehre in den Sozialwissenschaften. (Third edition). By O. ANDERSON. Würzburg, Germany: Physica-Verlag. 1957. Pp. 358. 24.50 DM.

This is the third edition of a book first published in 1954 in the excellent series of monographs issued by the German Statistical Society. As far as can be judged, it has become a standard 'intermediate' text in the German literature.

The scope of the book follows traditional lines. There are chapters on the basic ideas of statistical methods, on measures for describing frequency distributions, on index numbers, statistical errors, probability, time series, association, correlation and so forth. As the name of the book implies, its purpose is to highlight problems, rather than to give a comprehensive account of methods. This is all to the good and makes the book more interesting to read than the standard text-book. The level of approach is largely non-mathematical and should present few problems to anyone with ordinary school mathematics. Yet the book offers more than this implies, in that the basic derivations and mathematical background for the text are given in a 100-page mathematical appendix. Thus the interest of the book is extended to a wider and mathematically more sophisticated audience.

Certain aspects of order and balance are open to criticism. The placing of the chapters on index numbers and on statistical errors so early in the book is difficult to justify. More important, it seems to the reviewer that there should have been much greater and clearer emphasis on the basic concepts of sampling theory and sample design, and on problems of statistical inference as opposed to description. Conversely, the discussion of correlation (though not of regression) might have been shorter.

But these are matters of opinion which should not obscure the real value of Prof. Anderson's text. Students—for whom the book was originally intended—and social scientists generally can hardly fail to be stimulated by its lively exposition and by the many illustrations and analogies presented in the text. A good example of Prof. Anderson's approach at its best is the discussion of cost-of-living indices: in the course of some twenty pages virtually all the main problems of index number construction are discussed with reference to what is probably the most widely publicized and misunderstood statistic there is.

One other feature of the book deserves mention. Prof. Anderson has taken the opportunity of this new edition to bring his usage of technical terms into line with that recommended for the German section of Kendall and Buckland's *Dictionary of Statistical Terms*. This is an admirable step which other authors would do well to follow. Perhaps it is not too much to hope that a growing uniformity in the use of technical terms will be followed by some standardization of statistical notation.

C. A. MOSER

Demográfia, Vol. I, Part 1. Budapest: Központi Statisztikai Hivatal. 1958. Pp. 159. 12Ft.

This new journal is mainly in Hungarian with brief summaries of articles in Russian and English. The wide gulf between Hungarian and the languages with which one is more familiar implies that by no means all statistical tables are understood.

The first paper—A Few Problems of Hungary's Demographic Policy—includes an attempt to identify the influence of recent legislative measures—the 1953 ban on the use of birth control practice, its rescinding in 1956, and the legalizing of abortion. The author stresses the urgent need to reduce the apparently large number of abortions by encouraging the use of contraception—as yet not widely employed. It appears that the original paper includes details, omitted from the English Summary, which would prove interesting to many Western demographers.

Two other papers, on the measurement of infant mortality and population projections, rest heavily on recent and now familiar U.N. work in these fields. A fourth paper deals with the questions that should be asked and tabulations made in the 1960 Hungarian census.

Five shorter articles cover: Hungary's Vital Statistics in Recent Years (crude rates and trend in number and age structure of population); Industrial Commuting in Hungary; A study of 2000 cases examined by the National Institute for Sterility Research (abortion is found a major cause of secondary sterility); Preparation for the 1959 U.S.S.R. census, and Deaths from Heart Disease.

The impression is gained that in problems of methodology, the effort of overcoming the language difficulty will not be repaid, but the situation is quite different in relation to more sociological questions involving the Hungarian way of life.

N. H. CARRIER

American Housing and its Uses. By L. WINNICK. New York: John Wiley and Sons Inc. London: Chapman and Hall Ltd. 1957. Pp. 143. 44s.

This statistical study of density of occupation of rooms in the U.S.A. is one of the most interesting investigations in the housing field of recent years, and Mr Winnick's scholarly treatment inspires confidence in his conclusions. The main data used is that of the U.S. Census statistics of the numbers of persons per room for households of different sizes and houses of different types and sizes. Mr Winnick measures density of occupation, *space utilization* as he calls it, by the ratio of persons to rooms, the PPR ratio; while this is, he points out, an imperfect measure for many purposes, it is the best the data permits. Information is also available about house rents and values, types of tenure and household incomes, in sufficient detail for effective examination of the relation between variations in them and in the PPR ratios. Perhaps Mr Winnick's most important conclusion from this study is that size of household tends to be more important than either rent or income in determining variations in the PPR ratio, and he finds that in consequence the range of incomes is much greater than that of the PPR ratios. A multiple correlation analysis of data for a number of cities tends to confirm his conclusion. It is of particular interest that increases in the number of rooms per person and decreases in the PPR ratios have been on the average small compared to changes in real income since the beginning of this century, though, as Mr Winnick points out, this is entirely consistent with the other results. He suggests that the great relative increase in the cost of building as well as the emergence of new durable consumer goods such as motor cars, provides some explanation of the relatively slight decline in the number of persons per room since the beginning of the century.

It is impossible to give any indication here of the many other important conclusions Mr Winnick reaches. Their interest to housing experts in this country is enhanced by the fact that the general pattern of space utilization in the U.S. in terms of PPR closely resembles that in this country. The excellent use in the inquiry of data on tenures, rents and incomes to throw light on the pattern of demand for room space demonstrates the desirability of obtaining more information about them in this country through the Census of Population.

MARIAN BOWLEY

Genetics and the Improvement of Tropical Crops. An Inaugural Lecture by Sir JOSEPH HUTCHINSON. Cambridge University Press. 1958. Pp. 30. 3s. 6d.

All who attended Prof. Hutchinson's Inaugural Lecture felt it a memorable occasion, being an exposition, by a master, of the lessons which the recent era of plant breeding can bequeath to the next phase of the art. All improvement programmes hinge on the variability of the material initially available, and Hutchinson gives a penetrating discussion of the origin of variability in wild and crop plants. Variability is inversely related to selection pressure which depends on population size and ecology. Special importance attaches to the breeding system. Variability is lost in the spread of an inbred crop plant, but is maintained if outcrossing occurs with appreciable frequency. Systems of improvement depending on a limited number of inbred lines necessarily involve early loss of genetic potentiality. The way forward, particularly in providing local races for the underdeveloped countries, would appear to lie in the interbreeding bulk or in recurrent selection. In all plant breeding work, however, as Prof. Hutchinson indicates with telling examples, there are two constant requisites for success. The selection programme must keep to a clearly formulated aim. This necessitates methods of precise statistical control. The reader may expect that plant breeding will continue to be a progenitor of new biometrical techniques.

A. R. G. OWEN

Morphological Integration. By E. C. OLSON and R. L. MILLER. The University of Chicago Press, for whom Cambridge University Press act as agents. 1958. Pp. 317. 75s.

This book is a collection of material by two biologists, one a palaeontologist and one a marine geologist. Profs. Olson and Miller have suggested theoretical methods for the classification and correlation of biological data and, contrary to general practice, have tried out their theory on actual observations.

The reviewer is unable to judge of the worth of their theory from a biological point of view, but from a general scientific standpoint it is sensible and constructive. Broadly speaking the authors recognize that all physical measurements on any one biological unit are correlated. (It was the existence of such correlation which led Karl Pearson to develop the theory of multiple and partial correlation some fifty years ago.) Profs. Olson and Miller discuss regression and partial correlation models, and also models where the data are considered relative to their biological function. They thus get two types of measure, the ρ -type in which correlations, etc. are found and the F -type in which are measures associated according to their biological function.

Statisticians should read this book if only to get some idea of the 'old-fashioned' view of biometry which is just now becoming fashionable again. The wealth of data is also useful from many points of view. It is also refreshing to read a book which is not a text-book of statistical theory in disguise, but one in which the elementary statistical ideas play an important but secondary role in producing a new biological theory.

F. N. DAVID

Scientific Programming in Business and Industry. By ANDREW VAZSONYI. New York: John Wiley and Sons Inc. London: Chapman and Hall Ltd. 1958. Pp. xix + 474. 108s.

New books on mathematical techniques of operational research follow one another so closely, even from the same publishers, that it is natural to look for distinguishing features, concerning either topics, or treatment, or both. As to topics, the book under review contains in three parts (The Fundamentals, Mathematical Programming, Programming in Production and Inventory Control) 2 chapters introduction and illustration, 4 on Linear, 1 on Convex and 1 on Dynamic Programming, 1 on the Theory of Games, and 4 on Inventory Control and Scheduling. In University lectures, the author has covered 'a good part of the text in a six-month period of 45 h of lecturing' (p. ix).

As to presentation, operational research techniques can be treated as branches of applied mathematics, with theorems, proofs, corollaries, etc., or as parts of case studies, where they gain interest from their (often spurious) realism, and lose in generality.

The author claims that 'the method of presentation...is different from that of other mathematical books....There is not a single "proof" in this book. Instead of proving theorems a great deal of effort is spent in presenting arguments so that they make sense' (p. vi).

Much as the reviewer sympathizes with such an anti-purist aim, he does not feel that the problem has been solved here. 'There are a lot of equations in the book...because they are the best way to describe the particular business problem' (p. vi). Precisely. But having been presented with notation and functional relationships, the reader is then treated to lengthy verbal explanations, and the mathematics is wasted. Also, such an approach is in danger of slipping up: we are told on p. 284 that it was proven in Chapter V that the functionals of two dual problems (in Linear Programming) have the same numerical optima, but on p. 153, where the theorem is stated, we read: 'Proof of the theorem is omitted here.' It is doubtful whether either the business manager (for whom the book is written), or the mathematician (for whom, apparently, it is not), will be satisfied. As a matter of conjecture, it is the mathematician who is better served. He can in most cases guess the essence of the missing proof, and he can appreciate the great wealth of information in this well-written book. He will also be gratified at the extensive and amazingly rapid development of a subject, aimed at practical applications, but which must be admitted—grudgingly, or with approval—to be a branch of mathematics.

Sources are mentioned in the text, but there is no comprehensive bibliography. There is, however, an author and a subject-index. The make-up of the book is excellent, and only a few misprints have been found.

S. VAJDA

Introduction to Operations Research By C. WEST CHURCHMAN, RUSSELL L. ACKOFF and E. LEONARD ARNOFF. New York: John Wiley and Sons Inc.; London: Chapman and Hall Ltd. 1957. Pp. x + 645. 96s.

Introduction to Operations Research is one of the most ambitious books yet written in this field. It has been developed from the contents of an annual course in operational research which is presented at the Case Institute of Technology and it is intended to be of use both to prospective customers and practitioners. The former are recommended to read certain prescribed parts of the book, whilst the latter are advised to become familiar with the whole of it. Most of the illustrative examples are taken from

industrial problems since these are, on the whole, the most accessible, and usually command the greatest interest in the audiences for whom the book is primarily designed. A warning is given, however, that the reader should not therefore assume that operational research is peculiar to the industrial scene only.

The book is divided into ten parts of which the first three and the last deal with the general structure of operational research. The remaining parts, the filling of the sandwich, are concerned with a variety of special techniques which have proved themselves to be of great use to operational research workers.

The chapters on the structure of operational research are both the weakest and the most valuable part of the book. The ideas behind the design of these chapters are admirable, but unfortunately the authors have not been able to prevent some of the exposition being wordy and diffuse. Statements such as 'It is characteristic of operations research that in the solution of each problem new problems are uncovered', and the phases of an operational research problem which are listed on page thirteen could be applied with equal validity to many disciplines and illustrate the difficulty of pinpointing in a few words the essential nature of operational research itself. Chapter 2 'An Operations Research Study of a System as a Whole', was disappointing. Although the problem outlined on p. 21 seems to require that the output rate of the main product be reduced, the end of the first stage of the research left this rate unaltered, nor is there any mention in the second stage of the research of such a reduction or, alternatively, of a step-up in sales promotion in order to keep pace with production. The dangers of suboptimization are not discussed until much later in the book. The success of the market trend analysis described on pp. 49-52 is a resounding tribute to the robust character of the methods used and the conformable behaviour of the series under examination since the level of the Type I error must frequently have been higher than the classical value of 5%. Chapters 4 and 5 on the 'Analysis of the Organization and on the Formulation of the Problem', and the excellent case study of Chapter 3, are good and succeed in emphasizing the comprehensive approach required in operational research. The final chapter on the selection of a suitable team contains a useful discussion of the desirable attributes which members of the team should possess, but it is surprising that in a subject which is so much concerned with the maximization of profits and/or the minimization of costs there appears to be no place in the team for an economist.

In parts IV-VIII, five types of problem which are encountered time and again in operational research are discussed. These are problems of inventory control, of allocation, problems concerning queues and waiting lines, problems of the replacement of perishable material and problems of competition and strategy. Most statisticians will be acquainted with these problems, since a realistic formulation of any one of them is almost sure to be in terms of random variables. Although it can be argued that the lack of mathematical sophistication in the structural chapters is an advantage, it proves a drawback in this technical section. The authors' self-imposed limitations force them into lengthy discussion and a spate of footnotes. Further, they are, for the same reason, compelled to omit many of the more interesting cases of these problems. For example, most of part IV in which inventory control is considered, is entirely deterministic; the effect of adding restrictions to, or of altering some of the parameters in, a linear programming model is carried out only in a numerical example without any general rules of behaviour or conclusions being drawn; and although the section on queuing models discusses simple queues at some length, it barely mentions the problems of queues in series, of bulk service queues, and of queues of customers with different priorities for service, etc. The chapters on testing, control and implementation of solutions which comprise part IX should be prescribed reading for all operational research workers in order to remind them that their duties extend beyond the construction of a satisfactory mathematical model.

The book is well set-up and few misprints were noticed; the worst appeared to be the wrong page number quoted in the index for the all too short discussion of the dual theorem of linear programming. The index itself is rather brief. On the other hand, an excellent list of references is quoted at the end of each chapter.

Just as most operational research workers cannot agree on a definition of operational research, so also there are many opinions on the form which an introductory text-book should take. This is an operational research problem in itself to which further study should be given. *Introduction to Operations Research* will be of use to the intelligent, sceptical executive who wishes to survey the field, and probably should be read by all operational research workers, if only for the sake of part IX. But the authors have been unable to find a level of exposition which is acceptable both to consumer and to practitioner, and one is left with the impression that the filling could have been omitted from the sandwich without seriously affecting either its food value or its flavour.

ALISON DOIG

Principles of Statistical Techniques. A First Course from the beginnings for Schools and Universities. By P. G. MOORE. Cambridge: University Press. 1958. Pp. 239. 22s. 6d.

The appearance of this book is timely for those struggling with statistics in schools, with negligible direction from above in methods and application. The author has resisted the temptation to go too deeply into theory, and succeeds in making the book readable, stimulating and beneficial to anyone prepared to carry out the necessary calculations. His claim that 'this book is an attempt to put across the main principles of statistical methods to students, who are fundamentally interested in the practical applications of the subject...', is well founded. The thirty exercises at the end of the second chapter underline his attempt to cater for all tastes. The main asset of the book is the successful use of topics selected from many fields to illustrate principles and techniques. Dr Moore has gone some way towards providing a book which ought to stimulate teachers of many subjects into the realization of the advantages of a statistical approach.

Guidance in the collection and tabulation of data, so rarely given, is nicely handled, and the pitfalls of pictorial representation well illustrated. A genuine attempt to link up probability, sampling and the binomial theorem has been made, either as a natural consequence of events or as the starting point of a theory and once again the examples are plentiful. The emphasis, perhaps by bold type, of important ideas, principles, tests or formulae would have helped, and more summaries like that at the end of Chapter 10 would have been profitable. The omission of $\Sigma(O - E)^2/E$, where O, E are observed and expected frequencies in Chapter 11, is unfortunate, but economists should find the work on indices and time series in the next chapters to their benefit. Overrating the importance of correlation is given two shocks in the last chapter; first, by the title 'Pairs of Characters' and, secondly, by the excellent example in the last paragraph of the book.

This book deserves wide use in schools, both for reference and class use, but the cost would appear to be prohibitive for the latter. A set of answers to the numerical problems would also help.

A. P. PENFOLD

OTHER BOOKS RECEIVED

- The Advanced Theory of Statistics.** Vol. I. Distribution Theory. By M. G. KENDALL and A. STUART. London: Charles Griffin and Co. Ltd. 1958. Pp. xii + 433. 84s.
- Wahrscheinlichkeitsrechnung und Matematische Statistik.** (Hochschulbücher für Mathematik Band 40. By M. FISZ. Berlin: Veb Deutscher Verlag der Wissenschaften. 1958. Pp. 528. 14 DM.
- Planning of Experiments.** By D. R. COX. New York: John Wiley and Sons. Inc. London: Chapman and Hall Ltd. 1958. Pp. 308. 60s.
- Theory and Methods of Scaling.** By W. S. TORGERSON. New York: John Wiley and Sons Inc. London: Chapman and Hall Ltd. 1958. Pp. xi + 460. 76s.
- Linear Programming Methods and Applications.** By S. I. GASS. U.S.A. and London: McGraw-Hill Book Company. 1958. Pp. 223. 50s. 6d.
- Lectures on Ordinary Differential Equations.** By W. HUREWICZ. New York: John Wiley and Sons Inc. and The Technology Press of the Massachusetts Institute of Technology. 1958. Pp. xiii + 122. 40s.
- Introduction to Functional Analysis.** By A. TAYLOR. New York: John Wiley and Sons Inc. London: Chapman and Hall Ltd. 1958. Pp. 423. 100s.
- Surveys in Applied Mathematics.** New York: John Wiley and Sons Inc. London: Chapman and Hall Ltd. 1958. Vol. 1. Elasticity and Plasticity. By J. N. GOODIER and P. G. HODGE. Pp. 152. 50s. Vol. 2. Dynamics and Non linear Mechanics. By E. LEIMANIS and N. MINORSKY. Pp. 206. 62s. Vol. 3. Mathematical Aspects of Subsonic and Transonic Gas Dynamics. By L. BERS. Pp. 164. 62s. Vol. 4. Some Aspects of Analysis and Probability. By I. KAPLANSKY, M. HALL, E. HEWITT and R. FORTET. Pp. 243. 72s.
- The Structure of Arithmetic and Algebra.** By M. H. MARIA. New York: John Wiley and Sons Inc. London: Chapman and Hall Ltd. 1958. Pp. 284. 48s.
- Theorie des Graphes et ses Applications.** By C. BERGE. Paris: Dunod. 1958. Pp. 275.
- Introduccion a la Investigacion Operativa.** By SIXTO RIOS *et al.* Madrid: Nuevas Graficas SA, for Instituto de Investigaciones Estadisticas del Consejo Superior de Investigaciones Cientificas. Pp. 56.
- Tuberculosis in White and Negro Children.** Harvard: University Press for the Commonwealth Fund. London: Oxford University Press. 1958. Vol. I. The Roentgenologic Aspects of the Harriet Lane Study. By JANET B. HARDY. Pp. 122. 60s. Vol. II. The Epidemiological Aspects of the Harriet Lane Study. By MIRIAM E. BRAILEY. Pp. 103. 36s.
- Sampling Opinions.** An Analysis of Survey Procedures. By F. F. STEPHAN and P. J. MCCARTHY. New York: John Wiley and Sons Inc. London: Chapman and Hall Ltd. 1958. Pp. xvi + 451. 96s.
- Annual Epidemiological and Vital Statistics 1955.** (Bilingual edition English and French.) Geneva: World Health Organization. [Available from Her Majesty's Stationery Office.] 1958. Pp. 699. £3, \$12.00 or Sw.fr. 36.
- Contributo allo Studio delle Tavole di Nuzialità.** By G. PANIZZON. Padova, Italy: Cedam. 1958. Pp. 143.
- Table of Factors for One-sided Tolerance Limits for a Normal Distribution.** By DONALD B. OWEN. New Mexico: Sandia Corporation. 1958. Pp. 131.

- Table of Natural Logarithms for Arguments between Five and Ten to Sixteen Decimal Places.** Introduction by A. N. LOWAN. U.S. Dept. of Commerce Applied Maths. Series 53. 1958. Pp. xii + 506. \$4.00.
- Finite Queueing Tables.** (Publications in Operations Research, No. 2.) New York: John Wiley and Sons Inc. London: Chapman and Hall Ltd. 1958. Pp. xvi + 210. 68s.
- Table for the Solution of Cubic Equations.** By H. E. SALZER *et al.* U.S.A. and London: McGraw-Hill Book Company Inc. 1958. Pp. xv + 161. 58s.
- Table of Transport Integrals.** $J_n(x) \equiv \int_0^x e^z z^n (e^z - 1)^{-2} dz$. (N.B.S. Circular 595.) U.S. Dept. of Commerce. 1958. Pp. 46. 40 cents.
- Tables of Partitions.** (Royal Society Mathematical Tables Vol. 4). By H. GUPTA, C. E. GWYTHER and J. C. P. MILLER. Cambridge University Press for the Royal Society. 1958. Pp. 39 Introduction + Pp. 132 Tables. 63s.
- Tables for Bivariate Osculatory Interpolation over a Cartesian Grid.** By H. E. SALZER and GENEVIEVE M. KIMBRO. San Diego, California: Convair-Astronautics. 1958. Pp. 40.

CORRIGENDA

(i) *Biometrika* (1950), 37, p. 96

'On the statistical independence of quadratic forms in normal variates.' By A. C. Aitken.

line 30, read $AVBV$ (or $BVAV$) for AVB (or BVA).(ii) *Biometrika* (1958), 45, p. 91

'Some properties of runs in quality control procedures.' By P. G. Moore.

(a) In the denominator of equation (5)

read $(T+1-Tx)(1-\mu)x^{m+1}$ for $(T+1-\mu x)x^{m+1}$.

(b) In the headings of the table below equation (5)

read $1-q_m$ for q_m .(iii) *Biometrika* (1958), 45, p. 244

'Conditioned Markoff processes.' By W. A. O'N. Waugh.

line 6, read $p_i u_i / \sum_{j=0}^{\infty} p_j u_j$ for $p_i u_i \sum_{j=0}^{\infty} p_j u_j$.(iv) *Biometrika* (1958), 45, p. 327

'The properties of a stochastic model for two competing species.' By P. H. Leslie and J. C. Gower.

On the left-hand side of the first two members of equation (5.3)

read $L_1 \sigma_1^2$ for $L^2 \sigma_1^2$ and read $L_2 \sigma_2^2$ for $L^2 \sigma_2^2$, respectively.

General Notes drawn up for the Guidance of Authors on the Preparation of Manuscripts for *Biometrika*

Papers should be headed by an informative title, by the name or names of the author(s), and where appropriate by the name and address of the laboratory or institution where the work was performed. Manuscripts should normally be typed in double spacing with wide margins and on one side of the paper only; they should be on sheets of uniform size, numbered consecutively. In addition to the top copy it is editorially convenient, though not essential, that a second copy of the paper be submitted to the journal. It is helpful if a shortened version of the title, not exceeding 65 letters and spaces, or about eight words in length, could be supplied for a page-head.

Special care should be taken with mathematics, and any complicated mathematical expressions should be inserted very clearly in the manuscript. Where algebraic expressions are written in, particular attention should be given to distinguishing between capitals and lower-case letters, e.g. S , s , etc. Where Greek letters or unusual symbols are used it may help the printer if they are distinguished by writing or underlining in coloured ink, and a key provided for the compositor. Alternatively each unusual letter may be indicated by name in the margin at the first mention, e.g. ' μ = Greek mu', ' R_F = italic capital R and sub-script italic capital F '. Letters in script should also be indicated in the margin. Equation numbers should be placed after, not before, the equation.

Tables, except very small ones, should be on separate sheets of paper, since they are not set up at the same time as the body of the text; the position at which they are to be inserted should be indicated on the manuscript. They should have headings or footnotes which make their general meaning comprehensible to the reader without reference to the text.

Illustrations should be kept to a minimum. The author's name and the number of the figure should be written on the back of each, with legends on a separate sheet. Diagrams that are not drawn by professional draughtsmen generally have to be re-drawn and re-lettered before blocks are made, and for this reason it is advisable to draw illustrations carefully in pencil from which the printer will make good ink drawings. It is most important that no mistake should be made in any part of the author's original drawing.

The Harvard system is used for references in this journal, references being collected at the end of the paper in alphabetical order of author. Each reference must give name followed by initials of author, the year of publication in brackets, title of paper, journal title abbreviated in accordance with the *World List of Scientific Periodicals*, volume number in arabic numerals with wavy line to indicate black type and the number of the first and last page in arabic numerals.

Papers should be submitted in their final form so that proof correction is confined to a minimum.

BIOMETRIKA PUBLICATIONS

Issued by the Cambridge University Press, Bentley House, London, N.W. 1
and obtainable from any bookseller

Tables of the Incomplete B-Function

59 pages of Introduction and 494 pages of Tables

EDITED BY KARL PEARSON

Price: 55s. net

Tables of the Incomplete Γ -Function

31 pages of Introduction and 164 pages of Tables

EDITED BY KARL PEARSON

Price: 42s. net

Tables of the Complete and Incomplete Elliptic Integrals

(from LEGENDRE's *Traité des Fonctions Elliptiques*. With autographed portrait of LEGENDRE)
39 pages of Introduction by KARL PEARSON and 94 pages of Tables

Price: 12s. 6d. net

Tables of the Ordinates and Probability Integral of the Distribution of the Correlation Coefficient in Small Samples

38 pages of Introduction, 55 pages of Tables, 10 Diagrams and 4 Charts

By F. N. DAVID

Price: 17s. 6d. net

Biometrika Tables for Statisticians, Vol. I

EDITED BY E. S. PEARSON and H. O. HARTLEY for the Biometrika Trust
102 pages of Introduction and 136 pages of Tables

Price: 25s. net

The Life, Letters and Labours of Francis Galton, Vols. I, II, IIIA, & IIIB

By KARL PEARSON, F.R.S.

Price: £5. 5s. net

Karl Pearson: An Appreciation of Some Aspects of his Life and Work

By E. S. PEARSON

Price: 15s. net

A Bibliography of the Statistical and Other Writings of Karl Pearson

COMPILED BY G. M. MORANT, with the assistance of B. L. WELCH

Price: 6s. net

"Student's" Collected Papers

EDITED BY E. S. PEARSON and
JOHN WISHART with a FOREWORD BY LAUNCE MCMULLEN

Price: 21s. net

Karl Pearson's Early Statistical Papers

Reprinted by photo-lithography for the Biometrika Trust, with the permission of the original publishers. The Volume contains eleven papers, including the more important of the memoirs entitled "Mathematical Contributions to the Theory of Evolution", first published in the *Philosophical Transactions* of the Royal Society. The original paper deriving the χ^2 -distribution, published in 1900 in the *Philosophical Magazine*, is also included.

Price: 25s. net

PUBLICATIONS OF THE DEPARTMENT OF STATISTICS, UNIVERSITY COLLEGE, LONDON

*Issued by the Cambridge University Press, Bentley House, London, N.W. 1
and obtainable from any bookseller*

TRACTS FOR COMPUTERS

- I. **Tables of the Digamma and Trigamma Functions.** By ELEANOR PAIRMAN, M.A.
Tables for summing $S = \sum_{i=1}^{\infty} \frac{1}{(p_1 i + q_1)(p_2 i + q_2) \dots (p_n i + q_n)}$ where the p 's and q 's are numerical factors. *Price 5s. net.*
- V. **Table of Coefficients of Everett's Central-Difference Interpolation Formula.** By A. J. THOMPSON, PH.D. *Second edition. Price 7s. 6d. net.*
- VIII. **Table of the Logarithms of the Complete Γ -Function (to ten decimal places) for Argument 2 to 1200 beyond Legendre's Range (Argument 1 to 2).** By EGON S. PEARSON, D.SC. *Price 5s. net.*
- IX. **Log $\Gamma(x)$ from $x = 1$ to 50.9 by intervals of 0.01.** By JOHN BROWNLEE, M.D., D.SC. *Price 5s. net.*
- X. **On Quadrature and Cubature or on Methods of Determining Approximately Single and Double Integrals.** By J. O. IRWIN, D.SC. *Price 7s. 6d. net.*
- XII. **Tables of the Probable Error of the Coefficient of Correlation.** By KARL HOLZINGER, PH.D. *Price 5s. net.*
- XIII. **Bibliotheca Tabularum Mathematicarum, being a Descriptive Catalogue of Mathematical Tables. Part I. A, Logarithms of Numbers.** By JAMES HENDERSON, PH.D. *Price 9s. net.*
- XV. **Random Sampling Numbers.** By L. H. C. TIPPETT, M.SC., with a Foreword by KARL PEARSON. *Price 5s. net.*
- XXIII. **Tables of $\tan^{-1}x$ and $\log(1+x^2)$.** To assist in the calculation of the ordinates of a Pearson Type IV curve. By L. J. COMRIE, PH.D. *Price 5s. net.*
- XXIV. **Random Sampling Numbers (2nd Series).** By M. G. KENDALL and B. BABINGTON SMITH. *Price 5s. net.*
- XXV. **Random Normal Deviates.** By HERMAN WOLD. *Price 5s. net.*
- XXVI. **Correlated Random Normal Deviates.** By E. C. FIELLER, T. LEWIS and E. S. PEARSON. *Price 10s. 6d. net.*
- Nos. II, III, IV, VI and VII are out of print

LOGARITHMETICA BRITANNICA

A standard Table of Logarithms to Twenty Decimal Places. By A. J. THOMPSON, PH.D. (commenced in 1922 to commemorate the tercentenary of the publication of HENRY BRIGGS'S *Arithmetica Logarithmica*).

The nine separate sections of this Table have now been issued, and the complete work consisting of the logarithms of numbers 10,000–100,000, together with Dr Thompson's General Introduction (98 pp.) is available in two bound volumes.

Price £8. 8s. 0d.

NEW STATISTICAL TABLES: SEPARATES RE-ISSUED FROM BIOMETRIKA

To be obtained from
BIOMETRIKA OFFICE, UNIVERSITY COLLEGE, LONDON, W.C.1

- I. From *Biometrika*, Vols. 22, 27 and 28
Tests of Normality. By E. S. PEARSON and R. C. GEARY
Price 2s. 6d., post free
- II. From *Biometrika*, Vol. 32, pp. 168-181 and 188-189
(1) Table of percentage points of the incomplete beta-function
(2) Table of percentage points of the χ^2 distribution
Stitched together with introductory matter. Price 2s. 6d., post free
- III. From *Biometrika*, Vol. 32, pp. 300-310
(1) Table of the probability integral of the range in samples from a normal population
(2) Table of the percentage points of the range
(3) Table of the percentage points of the t-distribution
Stitched together with introductory matter. Price 2s. 6d., post free
- IV. From *Biometrika*, Vol. 33, pp. 73-88
Table of percentage points of the inverted beta (F) distribution
With introductory matter. Price 2s. 6d., post free
- V. From *Biometrika*, Vol. 33, pp. 252-265
(1) Table of the probability integral of the mean deviation in samples from a normal population
(2) Table of the percentage points of the mean deviation
Stitched together with introductory matter. Price 2s. 6d., post free
- VI. From *Biometrika*, Vol. 33, pp. 296-304
Table for testing the homogeneity of a set of estimated variances
With introductory matter. Price 2s., post free
- VII. From *Biometrika*, Vol. 35, pp. 145-156
Table of significance levels for the Fisher-Yates test of significance in 2×2 contingency tables. By D. J. FINNEY
With introductory matter. Price 2s. 6d., post free
- VIII. From *Biometrika*, Vol. 35, pp. 191-201
Table for the calculation of working probits and weights in probit analysis. By D. J. FINNEY and W. L. STEVENS
With introductory matter. Price 2s. 6d., post free
- IX. From *Biometrika*, Vol. 36, pp. 267-289
Tables of autoregressive series. By M. G. KENDALL
With introductory matter. Price 2s. 6d., post free
- X, XIV, XVIII, and XX. From *Biometrika*, Part 1 from Vol. 36, pp. 431-449, Parts 2 and 3 from Vol. 38, pp. 435-462, Part 4 from Vol. 40, pp. 427-446, and Part 5 from Vol. 42, pp. 223-242
Tables of symmetric functions. By F. N. DAVID and M. G. KENDALL
With introductory matter. Price 14s. 6d., post free (Part 1, 2s. 6d.; Parts 2 and 3, 4s.; Part 4, 4s.; Part 5, 4s.)
- XI. From *Biometrika*, Vol. 39, p. 190 and Vol. 43, pp. 449-451
Tables of percentage points of the extreme "Studentized" deviate from the sample mean. By K. R. NAIR and H. A. DAVID
With introductory matter. Price 1s., post free

NEW STATISTICAL TABLES: continued

- XII. From *Biometrika*, Vol. 37, pp. 168-172 and pp. 313-325
 (1) **Table of the probability integral of the t-distribution**
 (2) **Table of the χ^2 integral, and of the cumulative Poisson distribution.** By H. O. HARTLEY
 and E. S. PEARSON Stitched together with introductory matter. Price 5s., post free
- XIII. From *Biometrika*, Vol. 38, pp. 112-130
Charts of the power function for analysis of variance tests, derived from the non-central F-distribution. By E. S. PEARSON and H. O. HARTLEY
With introductory matter. Price 2s. 6d., post free
- XV. From *Biometrika*, Vol. 38, pp. 423-426
A chart for the incomplete beta-function and the cumulative binomial distribution. By H. O. HARTLEY and E. R. FITCH
With introductory matter and ruler scale. Price 2s. 6d., post free
- XVI. From *Biometrika*, Vol. 40, pp. 70-73
Tables of the angular transformation. By W. L. STEVENS
With introductory matter. Price 1s., post free
- XVII. From *Biometrika*, Vol. 40, pp. 74-86
Tests of significance in a 2×2 contingency table: extension of Finney's table (No. VII).
 Computed by R. LATSCHA With introductory matter. Price 2s. 6d., post free
- XIX. From *Biometrika*, Vol. 41, pp. 253-260
Tables of generalized k-statistics. By S. H. ABDEL-ATY With introductory matter. Price 2s., post free
- XXI. From *Biometrika*, Vol. 42, pp. 494-511
A new form of table for significance tests in a 2×2 contingency table. By P. ARMSSEN
With introductory matter. Price 2s. 6d., post free
- XXII. From *Biometrika*, Vol. 43, pp. 388-403
Tables for certain applications of sequential methods in the analysis of variance. By W. D. RAY
With introductory matter. Price 2s. 6d., post free
- XXIII. From *Biometrika*, Vol. 43, pp. 423-435
Table for determining confidence limits for a proportion in binomial sampling. By EDWIN L. CROW
With introductory matter. Price 2s. 6d., post free
- XXIV. From *Biometrika*, Vol. 44, pp. 411-419
Tables for estimating the normal distribution function of normit analysis. Part I. Tables and description of their use. By JOSEPH BERKSON
With introductory matter. Price 2s. 6d., post free
- XXV. From *Biometrika*, Vol. 44, pp. 482-489
Table of significance points for a two-sample t-test based on range. By P. G. MOORE
With introductory matter. Price 2s. 6d., post free
- XXVI. From *Biometrika*, Vols. 44 & 45
Tables of the upper percentage points of the generalized beta distribution. By F. G. FOSTER
 and D. H. REES With introductory matter. Price 5s., post free
- XXVII. From *Biometrika*, Vol. 46, pp. 178-204
Tables of 1000 standardized random deviates from certain non-normal distributions. By M. H. QUENOUILLE. With a note by E. S. Pearson
With introductory matter. Price 2s. 6d., post free

SEPARATES RE-ISSUED FROM BIOMETRIKA

To be obtained from

BIOMETRIKA OFFICE, UNIVERSITY COLLEGE, LONDON, W.C.1

The Biometrika Office has available for sale copies of many of the papers which have been published in *Biometrika* (right back to Volume 1). Readers requiring offprints of any papers are invited to enquire from the above address whether they are still available. The following are a few examples of fairly recent separates of which copies may be obtained:

From *Biometrika*, Vol. 39, pp. 324-345

Rank Analysis of Incomplete Block Designs I. The Method of Paired Comparisons. By R. A. BRADLEY and M. E. TERRY

Price 4s., post free

From *Biometrika*, Vol. 41, pp. 502-537

Rank Analysis of Incomplete Block Designs II. Additional Tables of Paired Comparisons. By R. A. BRADLEY

Price 5s., post free

From *Biometrika*, Vol. 43, pp. 203-205

Further critical values for the two-means problem. By W. H. TRICKETT, B. L. WELCH and G. S. JAMES

Price 1s., post free

From *Biometrika*, Vol. 44, pp. 1-8

John Wishart, 1898-1956. Obituary Notice and Bibliography. By E. S. PEARSON

Price 2s., post free

From *Biometrika*, Vol. 44, pp. 490-514

A bibliography on the theory of queues. By ALISON DOIG

Price 5s., post free

From *Biometrika*, Vol. 45, pp. 293-315

THOMAS BAYES'S Essay towards solving a problem in the doctrine of chances. [Reproduced from *Phil. Trans. Roy. Soc.* 1763, 53, 370-418.] With a biographical note by G. A. BARNARD.

Price 5s., post free

From *Biometrika*, Vol. 45, pp. 521-543

A bibliography on life testing and related topics. By WILLIAM MENDENHALL

Price 5s., post free

BIOMETRIKA INDEX. Comprising Subject Index for Vols. 1-37 and Author Index for Vols. 1-40, with Author Index Supplement covering Vols. 41-43.

Price 6s. or \$1.00, post free

STATISTICAL EXERCISES

Issued by the DEPARTMENT OF STATISTICS
UNIVERSITY COLLEGE, LONDON, W.C.1

Part I. Elementary Statistical Exercises.

Compiled by F. N. DAVID (91 pp.)

These exercises, published in 1953, were collected in connection with the numerical classwork undertaken by students during the first year of the B.Sc. Special Degree course in Statistics at University College. They deal with applications of the more elementary univariate and bivariate theory.

Price: 6s. 6d.

Part II. Statistical Exercises. Analysis of variance and associated techniques.

(107 pp.)

Compiled by N. L. JOHNSON

This volume includes over 100 exercises on the following topics:

Analysis of variance techniques for the following experimental designs: randomized block, Latin square, confounding, split plots, fractional replication. *Linear and curvilinear regression. Dosage-mortality techniques. Discriminant functions. Time series. Curve fitting.*

Price 12s.

[A supplementary volume containing suggested solutions to a substantial proportion of the exercises in Part II is in preparation.]

Both Parts are reproduced by offset-litho processes from typescript and bound in stiff paper covers. Copies may be obtained from a bookseller or direct from the Department of Statistics, University College, Gower Street, London, W.C.1. If ordered direct, payment must be made in advance and cheques made payable to University College London.

APPLIED STATISTICS

A JOURNAL OF THE ROYAL STATISTICAL SOCIETY

VOLUME VIII, No. 2

CONTENTS

JUNE 1959

Matching and Prediction on the Principle of Biological Classification. WILLIAM A. BELSON.
Deferred Sentencing Schemes. I. D. HILL, GARETH HORSNELL and BERNARD T. WARNER.
The Length of Cigarette Stubs. PERCY G. GRAY and ELIZABETH A. PARR.
The Use of Edge-Punched Cards in Statistical Computation. DENIS H. WARD.
Designing a Budget Survey. W. F. F. KEMSLEY.
Pseudo-Random Elements for Computers. E. S. PAGE.
LETTER TO THE EDITOR: Acceptance Sampling.
MEETINGS OF SECTIONS OF THE ROYAL STATISTICAL SOCIETY.
BOOK REVIEWS AND PUBLICATIONS RECEIVED.

NOVEMBER 1959

VOLUME VIII, No. 3

Adjusting Single Sampling Plans for Finite Lot Size. HUGO C. HAMAKER.
A Reliability Study of Sensory Assessments. A. S. C. EHRENBURG.
A Statistical Approach to Stores Auditing. J. S. JAMES.
Weekly, Monthly, and Quarterly Tolerances for Coke Quality. DENIS H. WARD.
A Problem of Subjective Classification in Industrial Medicine. J. R. ASHFORD.
Shipping Costs and the Terms of Trade: Australia and New Zealand. MICHAEL CHISHOLM.
NOTES AND COMMENTS.
MEETINGS OF SECTIONS OF THE ROYAL STATISTICAL SOCIETY.
BOOK REVIEWS AND PUBLICATIONS RECEIVED.

APPLIED STATISTICS is published three times per year. The annual subscription is £1. 10s. (U.S.A. and Canada \$5.00). Single copies 13s. post free. Orders should be sent to

Oliver and Boyd Ltd., Tweeddale Court, 14 High Street, Edinburgh, 1

TECHNOMETRICS

*A Journal of Statistics for the Physical,
Chemical and Engineering Sciences*

MAY 1959

CONTENTS

Measurements Made by Matching with Known Standards. W. J. YOUDEN, W. S. CONNOR and N. C. SEVERO.
Random Balance Experimentation. F. E. SATTERTHWAITHE.
The Application of Random Balance Designs. THOMAS A. BUDNE.
Discussion of the Papers of Messrs Satterthwaite and Budne. W. J. YOUDEN, O. KEMPTHORNE, J. W. TUKEY,
G. E. P. BOX and J. S. HUNTER.
Quick Analysis Methods for Random Balance Screening Experiments. F. J. ANSCOMBE.

AUGUST 1959

VOL. 1, No. 3

Simplified Estimators for the Normal Distribution when Samples are Single Censored or Truncated. A. CLIFFORD
COHEN, Jr.
Control Chart Tests Based on Geometric Moving Averages. S. W. ROBERTS.
The Measuring Process. JOHN MANDEL.
Factorial Experiments in Life Testing. MARVIN ZELEN.
The Use of LaGrange Multipliers with Response Surfaces. A. W. UMLAND and W. N. SMITH.
A Statistical Model for Evaluating the Reliability of Safety Systems for Plant Manufacturing Hazardous Products.
LOUIS B. KAHN.

Members of the American Statistical Association and the American Society for Quality Control may subscribe at a special rate of \$6.00 per year. Non-member subscriptions are \$8.00 per year. Remittances, made payable to Technometrics, may be sent to the office of either society, as follows:

AMERICAN STATISTICAL ASSOCIATION,
ROOM 404, BEACON BUILDINGS,
1757 K STREET, NW,
WASHINGTON 6, DC, U.S.A.

AMERICAN SOCIETY FOR QUALITY CONTROL,
6197 PLANKINTON BUILDING,
161 W. WISCONSIN AVENUE,
MILWAUKEE 3, WISCONSIN, U.S.A.

The Annals of Mathematical Statistics

The Official Journal of the Institute of Mathematical Statistics

VOL. 30, NO. 4

CONTENTS

DECEMBER, 1959

- Some Validity Criteria for Statistical Inferences. ROBERT J. BUEHLER.
Conditional Confidence Level Properties. DAVID L. WALLACE.
An Example of Wide Discrepancy Between Fiducial and Confidence Intervals. CHARLES STEIN.
Optimum Invariant Tests. E. L. LEHMANN.
The Weighted Compounding of Two Independent Significance Tests. M. ZELEN and L. S. JOEL.
Bayes Acceptance Sampling Procedures for Large Lots. D. GUTHRIE, Jr. and M. V. JOHNS, Jr.
Optimum Tolerance Regions and Power When Sampling from Some Non-Normal Universes. IRWIN GUTTMAN.
Properties of Model II—Type Analysis of Variance Tests, A: Optimum Nature of the F -Test for Model II in the Balanced Case. LEON H. HERBACH.
Some Remarks on Herbach's Paper, 'Optimum Nature of the F -Test for Model II in the Balanced Case'. WERNER GAUTSCHI.
The Most-Economical Character of Some Bechhofer and Sobel Decision Rules. WM. JACKSON HALL.
The Admissibility of Pitman's Estimator of a Single Location Parameter. CHARLES STEIN.
The Use of Sample Quasi-Ranges in Estimating Population Standard Deviation. H. LEON HARTER.
The Joint Cumulants of True Values and Errors of Measurement. FREDERIC M. LORD.
Some Tests of Permutation Symmetry. R. WORMLEIGHTON.
Contributions to the Theory of Rank Order Statistics—The One-Sample Case. I. RICHARD SAVAGE.
The Distribution of a Generalized D_n^+ Statistic. MEYER DWASS.
Null Distribution of the Hodges Bivariate Sign Test. JEROME KLOTZ.
Exact Nonparametric Tests for Randomized Blocks. JOHN E. WALSH.
A Generalization of Partially Balanced Incomplete Block Designs. B. V. SHAH.
The Non-Existence of Certain PBIB Designs. MANOHAR NARHAR VARTAK.
A Necessary Condition for Existence of Regular and Symmetrical Experimental Designs of Triangular Type, with Partially Balanced Incomplete Blocks. JUNJIRO OGAWA.
Optimal Spacing in Regression Analysis. H. A. DAVID and BEVERLY E. ARENS.
Third Order Rotatable Designs for Exploring Response Surfaces. D. A. GARDINER, A. H. E. GRANDAGE and R. J. HADER.
Second Order Rotatable Designs in Three Dimensions. R. C. BOSE and NORMAN R. DRAPER.
The Probability in the Extreme Tail of a Convolution. DAVID BLACKWELL and J. L. HODGES, Jr.
Bounds on Normal Approximations to Student's and the Chi-Square Distributions. DAVID L. WALLACE.
Approximate Expressions for the Conditional Mean and Variance over Small Intervals of a Continuous Distribution. GUNNAR EKMAN.
On the Moments of the Trace of a Matrix and Approximations to its Distribution. K. C. S. PILLAI and TITO A. MIJARES.
Random Graphs. E. N. GILBERT.
Scale Mixing of Symmetric Distributions with Zero Means. E. M. L. BEALE and C. L. MALLOWS.
Measurability of Extensions of Continuous Random Transforms. OTTO HANS.
A Convolution Class of Monotone Likelihood Ratio Families. S. G. GHURYE and DAVID L. WALLACE.
On the Laws of Cauchy and Gauss. R. G. LAHA.
Continuous Sampling Procedures without Control. C. DERMAN, M. V. JOHNS, Jr. and G. J. LIEBERMAN.
Some Convergence Theorems for Stationary Stochastic Processes. T. KAWATA.
Large Excursions of Gaussian Processes. MARK KAC and DAVID SLEPIAN.
The Capacity of a Class of Channels. DAVID BLACKWELL, LEO BREIMAN and A. J. THOMASIAN.
Infinite Codes for Memoryless Channels. DAVID BLACKWELL.

NOTES.

CORRECTION NOTES.

ABSTRACTS OF PAPERS.

PUBLICATIONS RECEIVED.

Subscription rate \$12.00 per year in the United States and Canada and \$10.00 per year elsewhere

ADDRESS ORDERS FOR SUBSCRIPTIONS AND BACK NUMBERS TO
PROFESSOR A. H. BOWKER, TREASURER, INSTITUTE OF MATHEMATICAL
STATISTICS, DEPARTMENT OF STATISTICS, STANFORD UNIVERSITY,
STANFORD, CALIFORNIA, U.S.A.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

VOL. 54, No. 287

CONTENTS

SEPTEMBER 1959

- NORMAN M. KAPLAN—Some Methodological Notes on the Deflation of Construction.
 WILLIAM A. CROMARTY—An Econometric Model for United States Agriculture.
 PETER E. DE JANOSI—A Note on the Relationship between Earning Expectations and New Car Purchases.
 CHARLES WINDLE—The Accuracy of Census Literacy Statistics in Iran.
 RONALD H. BEATTLE—Sources of Statistics on Crime and Correction.
 GORDON TULLOCK—Publication Decisions and Tests of Significance—A Comment.
 M. R. MICKEY—Some Finite Population Unbiased Ratio and Regression Estimators.
 OLIVE JEAN DUNN—Confidence Intervals for the Means of Dependent Normally Distributed Variables.
 G. E. P. BOX and NORMAN R. DRAPER—A Basis for the Selection of a Response Surface Design.
 JOHN W. PRATT—Remarks on Zeros and Ties in the Wilcoxon Signed Rank Procedure.
 S. I. ASKOVITZ—Graphic Methods Based upon Properties of Advancing Centroids.
 R. F. TATE and G. W. KLETT—Optimal Confidence Intervals for the Variance of a Normal Distribution.
 ENRICO T. FEDERIGHI—Extended Tables of the Percentage Points of Student's *t*-Distribution.
 G. P. STECK and D. B. OWEN—Percentage Points for the Distribution of Outgoing Quality.

For information about the Association and membership application forms, write
 THE SECRETARY, AMERICAN STATISTICAL ASSOCIATION,
 1757 K. ST., N.W., WASHINGTON 6, D.C., U.S.A.

ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

Contents of Vol. 28, No. 1, January 1960, include:

- ROBERT EISNER. A Distributed Lag Investment Function.
 GEORGE B. DANTZIG. On the Significance of Solving Linear Programming Problems with some Integer Variables.
 T. KLOEK and L. B. M. MENNES. Simultaneous-Equations Estimation based on Principal Components of Pre-determined Variables.
 H. S. HOUTHAKKER. The Capacity Method of Quadratic Programming.
 A. GHOSH. Input-Output Analysis with Substantially Independent Groups of Industries.
 A. G. DOIG and A. H. LAND. An Automatic Method of Solving Discrete Programming Problems.
 R. L. BASMANN. On the Asymptotic Distribution of an Infinite Class of Generalized Linear Estimators.
 W. V. CANDLER. A Short-Cut Method for the Complete Solution of Game Theory and Feed-Mix Problems.
 GEORGE H. BORTS. Estimation of Rail Cost Functions.
 MARC MERLOVE. The Market Demand for Durable Goods: A Comment.
 REPORT OF THE BILBAO MEETINGS.
 BOOK REVIEWS.
 ANNOUNCEMENTS AND NOTES.

Published Quarterly

Subscription rate available on request

The Econometric Society is an international society for the advancement of economic theory in its relation to statistics and mathematics.

Subscriptions to *Econometrica* and inquiries about the work of the Society and the procedure in applying for membership should be addressed to

RICHARD RUGGLES, Secretary
 THE ECONOMETRIC SOCIETY, BOX 1264, YALE UNIVERSITY
 NEW HAVEN, CONNECTICUT, U.S.A.

TRABAJOS DE ESTADISTICA

REVIEW PUBLISHED BY INSTITUTO DE INVESTIGACIONES ESTADÍSTICAS
OF THE CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

MADRID, SPAIN

Vol. X

CONTENTS

Cuaderno II

J. TALACKO—On stochastic linear inequalities.

D. E. BARTON and F. N. DAVID—A Collector's Problem.

F. AZORIN—Algunos problemas en la construcción de escalas de consumo.

NOTAS.

A. MARTIN ASIN—Análisis estadístico de las discordancias en una red de nivelación de alta precisión.

J. DE ARESPACOCCHAGA y R. DE MIGUEL—Estudio base para el plan de previsión estadística en la producción hidroeléctrica española.

CRONICAS.

BIBLIOGRAFIAS.

CUESTIONES Y EJERCICIOS.

For everything in connection with works, exchanges and subscriptions write to Professor Sixto Ríos, Instituto de Investigaciones Estadísticas, Serrano 123, Madrid, Spain.

The Review is composed of three fascicles published three times a year (about 350 pp.) and its annual price is 100 pesetas for Spain and \$4.00 U.S.A. for all other countries.

BIOMETRICS

Journal of the Biometric Society

VOL. 15, No. 3

CONTENTS

SEPTEMBER 1959

Biometric Method, Past, Present, and Future. J. O. IRWIN—Some Recent Results in Chi-Square Goodness-of-Fit Tests. G. S. WATSON—The Sampling Variance of the Genetic Correlation Coefficient. ALAN ROBERTSON—Confidence Limits for the LD_{50} Using the Moving Average-Angle Method. EUGENE K. HARRIS—The Use of the Power Function to Determine an Adequate Number of Progeny per Sire in a Genetic Experiment Involving Half-Sibs. STANLEY WEARDEN—Contribution to the Study of Grouped Observations. IV. Some Comments on Simple Estimates. N. F. GJEDDEBAEK—Optimum Group Size in Half-Sib Family Selection. J. M. RENDEL—Pair Comparison, With and Without Ties. N. T. GRIDGEMAN—On the Development of Clinical Statistical Systems for Psychiatry. J. B. CHASSAN—A Method for Testing Treatment Effects in the Presence of Learning. SEYMOUR GEISSER—The Comparison of the Sensitivities of Similar Experiments: Model II of the Analysis of Variance. D. E. W. SCHUMANN and R. A. BRADLEY—Test of Difference Between Treatment and Control with Multiple Replications of Control and a Missing Plot. DAVID HOGBEN.

VOL. 15, No. 4

DECEMBER 1959

Note on Optimum Group Size for Progeny Tests. A. W. NORDSKOG—Pleiotropism and the Genetic Variance and Covariance. C. J. MODE and H. F. ROBINSON—A Significantly Extreme Deviate in Data with a Non-Significant Heterogeneity χ^2 . W. F. BODMER—Combining Unbiased Estimators. FRANKLIN A. GRAYBILL and R. B. DEAL—An Illustration of the Use of Stochastic Approximation. LOUIS GUTTMAN and RUTH GUTTMAN—A Multiple Comparison Rank Sum Test: Treatments versus Control. ROBERT G. D. STEEL—An Approach to the Analysis of Data for Semi-Quantal Responses in Biological Assay. J. R. ASHFORD—Use of Contingency Tables in the Analysis of Consumer Preference Studies. R. L. ANDERSON—A Unified Theory for Quantal Responses to Mixtures of Drugs: Non-Interactive Action. P. S. HEWLETT and R. L. PLACKETT—Smallest Composite Designs for Quadratic Response Surfaces. H. O. HARTLEY—A Rapid Significance Test for Contingency Tables. MORTON KUPPERMAN—On a Quail Roadside Count Technique. EMIL H. JEBE.

'Biometrics' is published quarterly. The annual non-member subscription rate is \$7.00. Subscriptions, inquiries, and orders for back issues should be sent to:

BIOMETRICS
DEPARTMENT OF STATISTICS, VIRGINIA POLYTECHNIC INSTITUTE
BLACKSBURG, VIRGINIA, U.S.A.

Formerly ANNALS OF EUGENICS

October 1959

CONTENTS

October 1959

CONTENTS

Vol. 23, Pt. 4

Re-examination of a family with Pick's disease. V. W. D. SCHENK.—A contribution to the study of birthweight and survival of twins based on an Indian sample. N. K. NAMBOODIRI and V. BALAKRISHNAN.—The haptoglobins and hemoglobins of Alaskan Eskimos and Indians. B. S. BLUMBERG, A. C. ALLISON and BARBARA GARRY.—A note on deaf-mutism. C. S. CHUNG, O. W. ROBINSON and N. E. MORTON.—Further observations on the results of pregnancy-hyperoxaluria. E. F. SCOWEN, R. W. E. WATTS and E. G. HALL.—Observations on the results of pregnancies in women resident in Belfast: I. Data relating to all pregnancies ending in 1957. A. C. STEVENSON and H. A. WARNOCK.—II. Abortions or hydatidiform moles and ectopic pregnancies. A. C. STEVENSON, M. Y. DUDGEON and H. I. MCCLURE.—III. Sex ratio with particular reference to nuclear sexing of chorionic villi of abortions. A. C. STEVENSON.—A genetical study of ethanalamine phosphate excretion in hypophosphatasia. H. HARRIS and ELIZABETH B. ROBSON.—Cystathioninuria. H. HARRIS, L. S. PENROSE and D. H. THOMAS.—The effect of parental age, parity and duration of marriage on the twinning rate. M. C. BULMER.—The correlations between ridge-counts on different fingers estimated from a population sample. SARAH B. HOLT.—REVIEWS.

February 1960

February 1960

ridge-counts on different fingers estimated from a population. V. BALA-
Vol. 24, Pt. 1
The familial incidence of twinning. M. G. BULMER.—Data on the gestation period of Indian mothers. V. BALA-
KRISHNAN and N. K. NAMBOODIRI.—Micro-evolution among the Susak Islanders. Inbreeding, sterility, blood
groups and red hair. ZLATA DOLINAR.—The intellectual resemblance between sibs. J. MAXWELL and A. E. G.
PILLNER.—A genetical analysis of thirty families with gonadal dysgenesis. M. FRACCARO, K. KAIJSE
complement in continuously cultured cells of two individuals with gonadal dysgenesis. M. FRACCARO, K. KAIJSE
and J. LINDSTEN.—The occurrence of two unusual serum protein phenotypes in a single pedigree. H. HARRIS,
SYLVIA D. LAWLER, ELIZABETH B. ROBSON and O. SMITHIES.—Contribution to the heredity of the P.T.C. taste
character. JOSÉ PONS.—A note on the frequency of ABO heterospecific pregnancy. G. KNOX.—The chromosomes
in two patients with Klinefelter Syndrome. S. BERGMAN, J. REITALU, H. NOWAKOWSKI and W. LENZ.—Ectrodactyly.
Evidence in favour of a disturbed segregation in the offspring of affected males. A. C. STEVENSON and L. M.
JENNINGS.—Taste sensitivity to PTC in samples from three Brazilian populations. A. FREIRE-MAIA and A. QUELCE-
SALGADO.—REVIEWS.

Subscription price 100s. net per volume of four quarterly parts (in U.S.A. \$17.50) post.

CAMBRIDGE UNIVERSITY PRESS
BENTLEY HOUSE, 200 EUSTON ROAD, LONDON, N.W.1

SANKHYA
THE INDIAN JOURNAL OF STATISTICS
P. C. MAHALANOBIS

INDIAN JOURNAL OF
EDITED BY P. C. MAHALANOBIS
CONTENTS

CONTENTS

EDITED BY P. C. U. S. CONTENTS
Vol. 21, parts 3 and 4, 1959
 The analysis of heterogeneity. I. J. B. S. HALDANE—Sufficient statistics of minimal dimension. EDWARD W. BARANKIN and MELVIN KATZ, Jr.—The family of ancillary statistics. D. BASU—Metricizing rank-ordered or un-ordered data for a linear factor analysis. LOUIS GUTTMAN—Positive and negative dependence of two random variables. H. S. KONIJN—Definition and use of generalized percentage points. JOHN E. WALSH—Joint asymptotic distribution of U-statistics and order statistics. J. SETHURAMAN and B. V. SUKHATME—Some sampling systems providing unbiased ratio estimators. N. S. NANJAMMA, M. N. MURTHY and V. K. SETHI—Tables for some small sample tests of significance for Poisson distributions and 2×3 contingency tables. I. M. CHAKRAVARTI and C. RADHAKRISHNA RAO—Expected values of mean squares in the analysis of incomplete block experiments and fractional factorial designs with confounding and fractional replication. C. RADHAKRISHNA RAO—Some remarks on the missing plot analysis. SUJIT KUMAR MITRA—The use of linear algebra in deriving prime power factorial designs with confounding and fractional replication. NORMAN T. J. BAILEY—Properties of the invariant I_m (m -odd) for distributions admitting sufficient statistics. B. RAJA RAO—Numerical evaluation of a multivariate normal distribution. S. JOHN—On the evaluation of the probability integral of a multivariate normal distribution. S. JOHN—The distribution of Wald's classification statistic when the dispersion matrix is known. S. JOHN—An extension of Hald's table for the one-sided censored normal distribution. NIKHILESH BHATTACHARYA—Almost unbiased ratio estimates based on interpenetrating sub-sample estimates. M. N. MURTHY and N. S. NANJAMMA—Precision in the construction of cost of living index numbers. K. S. BANERJEE—Price indexes and sampling. ERLAND V. HOFSTEN—Corrigenda: A. SREE RAMA SASTRY; SAIBAL KUMAR BANERJEE.

CURRENT	per volume Rs. 45	per issue Rs. 12/8 \$4.50
---------	----------------------	---------------------------------

INDIA
FOREIGN

per volume
Rs. 30
\$10.00

per issue
Rs. 10
\$3.50

BACK NUMBERS

per volume
Rs. 45
\$15.00

per issue
Rs. 12/8
\$4.50

per volume
Rs. 30
\$10.00

Subscriptions and orders for back numbers should be sent to
204/1 Barrackpore Trunk R

INDIA
FOREIGN

RS. 30
\$10.00

Subscriptions and orders for back numbers should be sent to
Statistical Publishing Society, 204/1 Barrackpore Trunk Road, Calcutta-35

ROYAL STATISTICAL SOCIETY

THE JOURNAL OF THE ROYAL STATISTICAL SOCIETY is published in two series: SERIES A (GENERAL), four issues a year, annual subscription £4. 4s. post free; SERIES B (METHODOLOGICAL), two issues a year, annual subscription £3. 2s. post free.

SERIES A (GENERAL), VOL. 122, PART 2, 1959

- The Determinants of Wage Inflation: United Kingdom, 1946-56. By L. A. DICKS-MIREAUX and J. C. R. DOW. (With Discussion.)
- Production and Productivity Movements in the United Kingdom since 1900. By K. S. LOMAX. (With Discussion.)
- Using an Electronic Computer in a Problem of Medical Diagnosis. By T. H. HOLLINGSWORTH.
- Note on the Measurement and Prediction of Labour Turnover. By D. J. BARTHOLOMEW.
- Proceedings of a Special General Meeting held on 18 December, 1958.
- Obituary: F. A. Burchardt, 1902-1958; Arthur Cecil Pigou, 1877-1959.
- Reviews of Books, Statistical Notes, Additions to Library.

SERIES B (METHODOLOGICAL), VOL. 21, NO. 2, 1959

- Control Charts and Stochastic Processes. By G. A. BARNARD. (With Discussion.)
- Optimum Experimental Designs. By J. KIEFER. (With Discussion.)
- A Contribution to the Theory of Bulk Queues. By RUPERT G. MILLER, Jr.
- Maximising a Function in a Convex Region. By G. ZOUTENDIJK.
- Estimation of Location and Scale Parameters for the Rectangular Population from Censored Samples. By A. E. SARHAN and B. G. GREENBERG.
- The Distribution of Vacancies on a Line. By E. S. PAGE.
- The Output Process of the Queueing System M/G/1. By P. D. FINCH.
- The Efficiency of Automatic Winding Machines with Constant Patrolling Time. By A. J. HOWIE and L. R. SHENTON.
- The Comparison of Regression Variables. By E. J. WILLIAMS.
- The Most Economical Sequential Sampling Scheme for Inspection. By G. B. WETHERILL.
- Note on Fisher's Transformation of the Correlation Coefficient. By N. F. LAUBSCHER.
- The Analysis of Experimentally Distributed Life-Times with Two Types of Failure. By D. R. COX.

ROYAL STATISTICAL SOCIETY, 21 BENTINCK STREET, LONDON, W.1

AN INVESTIGATION OF HARTLEY'S METHOD FOR FITTING AN EXPONENTIAL CURVE

BY H. D. PATTERSON AND S. LIPTON

Rothamsted Experimental Station

INTRODUCTION

Hartley (1948) showed that the parameters of an exponential regression curve

$$\mathcal{E}(y_x) = \alpha - \beta \rho^x \quad (0 < \rho < 1), \quad (1)$$

where $x = 0, 1, 2, \dots, n-1$, can be estimated by considering the regression of y_x on x and on certain partial sums of the y_x . Since then, Stevens (1951) has described an iterative method for obtaining the least-squares estimates of α , β and ρ which is very suitable for routine applications when a high-speed computer is available and which, moreover, can be used when the x 's are not equally spaced. When computing facilities are limited, however, Stevens's method is somewhat laborious. For this reason other methods, including Hartley's, are still in common use.

Hartley (1948) showed that his estimates of ρ are of high efficiency for very large n in the special case $\alpha = 0$. It might reasonably be supposed that the estimates are also highly efficient for finite n and for α unknown although this does not necessarily follow. The purpose of the present paper is to discuss the efficiency and bias of these estimates and of other estimates of similar type using general formulae derived by Patterson (1958).

We will not be concerned with estimates of α and β as these can always be obtained by the linear regressions of y_x on r^x where r is an estimate of ρ . If r is subject to large errors or magnitude of $r - \hat{r}$, where $\hat{r}$ is the least-squares estimate. If r is subject to large errors or biases the fit may be very poor. On the other hand, if the errors and biases in r are small and the true relationship is in the form of (1) a good fit can always be obtained.

LINEAR AND QUADRATIC ESTIMATES OF ρ

Provided that $\sum_{x=1}^{n-1} w_x = 0$ the expression

$$r = \frac{\sum_{x=1}^{n-1} w_x y_x}{\sum_{x=1}^{n-1} w_x y_{x-1}} = \frac{A_1}{A_2} \quad (2)$$

provides consistent estimates of ρ . Several estimates of this type have been proposed (though not necessarily in this form). These include:

- (1) Stevens's least-squares estimate $\hat{r}$ in which the w_x are complicated functions of $\hat{r}$ itself.
- (2) Estimates in which A_1 and A_2 are both linear functions of the y_x . These estimates will be called 'linear estimates'. Suitable linear estimates with the w_x not depending on ρ have been discussed by Patterson (1956) for the cases $n = 4, 5, 6$ and 7 .
- (3) Estimates in which the w_x are linear functions of the y_x so that A_1 and A_2 are quadratic functions of the y_x . These estimates will be described as 'quadratic estimates'. As will be shown in the next section Hartley's estimate of ρ is a quadratic estimate.

It is possible, by a suitable choice of the w_x , to obtain both linear and quadratic estimates having minimum variance for some particular value of ρ , say ρ_0 . The quadratic estimates with this property cover a wider range of ρ (around ρ_0) with high efficiency than do the corresponding linear estimates but, as pointed out by Patterson (1958), they may be subject to large biases in some circumstances.

Before Hartley's estimates of ρ are considered in detail some properties of linear estimates, particularly those having minimum variance, will be presented. The variances and expectations of Hartley's estimates will then be determined for comparison with the minimum variance linear estimates. In the following it will be assumed that the y_x are independently and normally distributed with variance σ^2 .

The expectation and variance of a linear estimate of r are approximately

$$\mathcal{E}(r) = \rho + \left\{ \frac{\rho \sum_1^{n-1} w_x^2 - \sum_1^{n-2} w_x w_{x+1}}{\left(\sum_1^{n-1} w_x \rho^{x-1} \right)^2} \right\} \frac{\sigma^2}{\beta^2} \quad (3)$$

and

$$\text{var}(r) = \left\{ \frac{(1 + \rho^2) \sum_1^{n-1} w_x^2 - 2\rho \sum_1^{n-2} w_x w_{x+1}}{\left(\sum_1^{n-1} w_x \rho^{x-1} \right)^2} \right\} \frac{\sigma^2}{\beta^2}. \quad (4)$$

These expressions are suitable if σ^2 is small relative to A_2^2 .

Patterson (1958) has shown that for $\rho < 1$ the expression for $\text{var } r$ is at a minimum when $\rho = \rho_0$ if

$$(1 + \rho_0^2) w_x - \rho_0 (w_{x+1} + w_{x-1}) = k_1 (\rho_0^{x-1} + k_2), \quad (5)$$

where $x = 1, 2, \dots, n-1$; $w_0 = w_n = 0$; k_1 is given any convenient value such that the w_x are not all zero and k_2 is such that $\sum w_x$ is zero. A solution of equation (5) (with $k_1 = 1$) is

$$w_x = \sum_j d_{ij} \rho_0^{j-1} \quad (x = i), \quad (6)$$

where

$$d_{ij} = c_{ij} - \frac{\sum_j c_{ij} \sum_i c_{ij}}{\sum_{i,j} c_{ij}}, \quad (7)$$

and

$$c_{ij} = c_{ji} = \frac{\rho_0^{j-i} (1 - \rho_0^{2i}) (1 - \rho_0^{2(n-j)})}{(1 - \rho_0^{2n}) (1 - \rho_0^2)} \quad (i \leq j). \quad (8)$$

Equation (5) shows that, in this case (i.e. $k_1 = 1$),

$$(1 + \rho^2) \sum_1^{n-1} w_x^2 - 2\rho \sum_1^{n-2} w_x w_{x+1} = \sum_1^{n-1} w_x \rho^{x-1} \quad (9)$$

when $\rho = \rho_0$. Consequently, the variance of the minimum variance estimate is

$$\text{var}(r) = \left\{ \frac{1}{\sum_1^{n-1} w_x \rho^{x-1}} \right\} \frac{\sigma^2}{\beta^2}, \quad (10)$$

where the w_x are as defined in (6).

Linear estimates having full efficiency when $\rho = \rho_0$ can conveniently be denoted by $r(\rho_0)$. It is also possible to obtain estimates which tend towards full efficiency as ρ tends to 1.

These estimates will be denoted by $r(1)$. As ρ_0 approaches 1 the w_x given by (6) tend to be proportional to $\sum_j j d_{ij}$ where $x = i$. When $\rho_0 = 1$

$$c_{ij} = c_{ji} = i(n-j)/n \quad (i \leq j). \quad (11)$$

Consequently $r(1)$ is given by equation (2) with

$$w_x = \frac{1}{12}x(n-x)(2x-n). \quad (12)$$

As pointed out by Patterson (1956) the linear estimates $r(\rho_0)$ are equal to $\hat{r}$, the least-squares estimates of ρ , when $\rho_0 = \hat{r}$. The approximate formulae given above for the expectation and variance of $r(\rho)$ are also appropriate for use with the least-squares estimate $\hat{r}$.

Quadratic estimates of ρ with minimum variance when $\rho = \rho_0$ are given by the ratio (3) with

$$w_x = \sum d_{ij}(ky_{i-1} + ly_i), \quad (13)$$

where $x = i, j = 1, 2, \dots, n-1$, at least one of k and l is non-zero and the d_{ij} are as defined in (7) and (8) for $\rho_0 < 1$ or (7) and (11) for $\rho_0 = 1$. Only the value of ρ_0 and the ratio k/l are required to specify these estimates. Thus, the k and l of (13) can be multiplied by any non-zero finite number without affecting the estimates of ρ . A convenient notation for quadratic estimates with minimum variance when $\rho = \rho_0$ is, therefore, $r(\rho_0, k/l)$. This differs slightly from the notation previously used by Patterson (1958).

The expectation and variance of $r(\rho_0, k/l)$ are approximately

$$\mathcal{E}(r) = \rho + \frac{\sigma^2}{\beta^2} \left\{ \frac{(l - \rho k) \text{trace } \mathbf{D} + (k - \rho l) \text{trace } \mathbf{D}\mathbf{U}}{(k + l\rho) F_0} \right\} + \frac{\sigma^2}{\beta^2} \left\{ \frac{(l\rho^2 + 2k\rho - l) F_1 - 2kF_2}{(k + l\rho) F_0^2} \right\} \quad (14)$$

$$\text{var}(r) = \frac{\sigma^2}{\beta^2} \left\{ \frac{(1 + \rho^2) F_1 - 2\rho F_2}{F_0^2} \right\}, \quad (15)$$

and

$$F_0 = \sum_0^{n-1} W_x \rho^{x-1}, \quad F_1 = \sum_0^{n-1} W_x^2, \quad F_2 = \sum_0^{n-2} W_x W_{x-1}, \quad (16)$$

where

$$W_x = \sum_j d_{ij} \rho^{j-1}, \quad x = i, \quad (17)$$

$\mathbf{D}$ is the matrix of d_{ij} defined by equations (7) and (8) or (7) and (11) and $\mathbf{U}$ is an $(n-1) \times (n-1)$ matrix

$$\mathbf{U} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}. \quad (18)$$

The derivation of these expressions has been given by Patterson (1958). Thus the variance of $r(\rho_0, k/l)$ is independent of k/l . When $\rho = \rho_0$ (10) and (15) are identical. The family of estimates $r(1, k/l)$ include Hartley's estimate. The

As will be shown later the family of estimates $r(1, k/l)$ will therefore be considered in rather more detail. expectation and variance of $r(1, k/l)$ will therefore be considered in rather more detail. (19)

For these estimates the d_{ij} are given by (7) and (11) and

$$\text{trace } \mathbf{D} = \frac{1}{15}(n^2 - 4), \quad \text{trace } \mathbf{D}\mathbf{U} = \frac{1}{30}(2n^2 - 15n + 22). \quad (20)$$

For the purpose of studying the bias and variance of $r(1, k/l)$ when n is very large it is convenient to express the W_x in the alternative form

$$W_x = a + bx + cx^2 + d\rho^x,$$

where $d = -1/(1-\rho)^2$ and a, b and c are such that $W_0 = W_n = \Sigma W_x = 0$. For large n , constant $\theta = \rho^{n-1}$ and $x = nz$

$$W_x \rightarrow \frac{n^2}{(\ln \theta)^2} (1 + b'z + c'z^2 - \theta z), \quad (21)$$

where

$$b' = -2 \left\{ 2 + \theta + \frac{3(1-\theta)}{\ln \theta} \right\}, \quad (22)$$

$$b' + c' = -1 + \theta. \quad (23)$$

If

$$L_1 = \left(-1 + \frac{b'^2}{3} + \frac{c'^2}{5} + \frac{b'c'}{2} \right) - \frac{(1-\theta)(3-\theta)}{2 \ln \theta} \quad (24)$$

and

$$L_2 = -\frac{(1-\theta^2)}{2 \ln \theta} - \frac{\frac{1}{3}(1-\theta)^2 + c'^2}{(\ln \theta)^2}, \quad (25)$$

$$\beta^2 n^3 \text{var}(r)/\sigma^2 \rightarrow (\ln \theta)^2 (L_1 - L_2)/L_2^2. \quad (26)$$

It can also be shown, using the technique described by Stevens (1951, pp. 264-5), that $n^3 \text{var} r(\rho)$ tends to limiting values for constant θ as n is increased. Hence the efficiencies of $r(1, k/l)$ for very large n depend only on θ .

The bias in $r(1, k/l)$, when $k+l \neq 0$, tends to

$$\frac{\sigma^2}{n^2 \beta^2} \left[\frac{(\ln \theta)^2 \{15(l-k) - 2(k+l) \ln \theta\}}{30(k+l)L_2} + \frac{2 \ln \theta (L_1 - 2L_2)}{L_2^2} \right] \quad (27)$$

as n increases with θ constant.

HARTLEY'S ESTIMATE OF ρ

Hartley (1948)[†] noted that, if $y = \alpha - \beta \rho^x$, then

$$y_x = b_1 + b_2 x + b_3 Y_x, \quad (28)$$

where

$$Y_x = Y_{x-1} + \frac{1}{2} y_{x-1} + \frac{1}{2} y_x, \quad (29)$$

$$\left. \begin{aligned} 2Y_0 &= -y_0 - 2(y_1 + y_2 + \dots + y_{\frac{1}{2}(n-2)}) && \text{for even } n \\ &= -y_0 - 2(y_1 + y_2 + \dots + y_{\frac{1}{2}(n-3)}) - y_{\frac{1}{2}(n-1)} && \text{for odd } n \end{aligned} \right\} \quad (30)$$

and

$$\rho = (2 + b_3)/(2 - b_3).$$

He therefore suggested that α, β and ρ could be estimated by fitting (28) to the data by least squares. He called this procedure 'internal regression'.

The observational equations can be written in matrix notation as

$$\mathbf{y} = \mathbf{1}_n b_1 + \mathbf{x} b_2 + \mathbf{Y} b_3, \quad (31)$$

where $\mathbf{y}$ is a column vector with elements y_x , $\mathbf{1}_n$ is a column vector consisting of n 1's and $\mathbf{x}$ is a column vector with elements $x = 0, 1, 2, \dots, n-1$. The least-squares normal equations are then

$$\left. \begin{aligned} \mathbf{1}_n' \mathbf{y} &= n b_1 + \mathbf{1}_n' \mathbf{x} b_2 + \mathbf{1}_n' \mathbf{Y} b_3, \\ \mathbf{x}' \mathbf{y} &= \mathbf{x}' \mathbf{1}_n b_1 + \mathbf{x}' \mathbf{x} b_2 + \mathbf{x}' \mathbf{Y} b_3, \\ \mathbf{Y}' \mathbf{y} &= \mathbf{Y}' \mathbf{1}_n b_1 + \mathbf{Y}' \mathbf{x} b_2 + \mathbf{Y}' \mathbf{Y} b_3. \end{aligned} \right\} \quad (32)$$

[[†] See also, Hartley's note printed on p. 293 below. Ed.]

For the purpose of obtaining an explicit expression for b_3 (and hence for the estimate of ρ) it is convenient to introduce the matrices

$$\mathbf{T} \text{ (of order } n-1 \times n) = \begin{pmatrix} -1 & 1 & 0 & 0 & \dots \\ 0 & -1 & 1 & 0 & \dots \\ 0 & 0 & -1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \dots \end{pmatrix},$$

$$\mathbf{V} = \mathbf{T}\mathbf{T}',$$

$$\mathbf{D}_1 = \mathbf{V}^{-1} - \frac{\mathbf{V}^{-1}\mathbf{1}\mathbf{1}'\mathbf{V}^{-1}}{\mathbf{1}'\mathbf{V}^{-1}\mathbf{1}},$$

$\mathbf{J}$, an $n \times n$ matrix with diagonal terms equal to $(n-1)/n$ and non-diagonal terms equal to $-1/n$, and $\mathbf{1}$, a column vector with $n-1$ elements each equal to 1. It can be verified that

$$\mathbf{T}'\mathbf{V}^{-1}\mathbf{T} = \mathbf{J}$$

and that the elements d_{ij} of $\mathbf{D}$ are as defined by equations (7) and (11).

Elimination of b_1 from (32) yields the two equations:

$$\left. \begin{aligned} \mathbf{x}'\mathbf{J}\mathbf{y} &= \mathbf{x}'\mathbf{J}\mathbf{x}b_2 + \mathbf{x}'\mathbf{J}\mathbf{y}b_3, \\ \mathbf{X}'\mathbf{J}\mathbf{y} &= \mathbf{Y}'\mathbf{J}\mathbf{x}b_2 + \mathbf{Y}'\mathbf{J}\mathbf{y}b_3. \end{aligned} \right\} \quad (33)$$

Since $\mathbf{x}'\mathbf{T}' = \mathbf{1}'$, $\mathbf{T}\mathbf{y} = \mathbf{y}_1 - \mathbf{y}_0$ and $2\mathbf{T}\mathbf{y} = \mathbf{y}_0 + \mathbf{y}_1$, where $\mathbf{y}_0$ is a column vector $(y_0, y_1, \dots, y_{n-2})$ and $\mathbf{y}_1$ is a column vector $(y_1, y_2, \dots, y_{n-1})$, the regression b_3 can be written

$$b_3 = \frac{2(\mathbf{y}_0' + \mathbf{y}_1')\mathbf{D}_1(\mathbf{y}_1 - \mathbf{y}_0)}{(\mathbf{y}_0' + \mathbf{y}_1')\mathbf{D}_1(\mathbf{y}_0 + \mathbf{y}_1)}.$$

Hence the estimate of ρ is

$$\frac{2 + b_3}{2 - b_3} = \frac{(\mathbf{y}_0' + \mathbf{y}_1')\mathbf{D}_1\mathbf{y}_1}{(\mathbf{y}_0' + \mathbf{y}_1')\mathbf{D}_1\mathbf{y}_0}. \quad (34)$$

This is equivalent to equation (2) with the w_x given by (13), $\rho_0 = 1$, and $k/l = 1$. Consequently, Hartley's estimate of ρ is equal to $r(1, 1)$, one of the family of quadratic estimates which tend towards full efficiency as ρ tends to 1.

It will be noted that the estimation of ρ does not depend on the definition of Y_0 . Nor does the estimate of α , given by $-b_2/b_3$. Alternative estimates of β can, however, be obtained depending on Y_0 . With Y_0 given by equation (30), $\beta\rho^{\frac{1}{2}(n-1)}$ is estimated by

$$(-b_2/b_3) - b_1 - (n-1)b_2/2.$$

If Y_0 is taken to be zero the estimate of β is $-[b_2/b_3] - b_1$.

ALTERNATIVE INTERNAL REGRESSION METHODS

The reasons behind Hartley's choice of equations (28), (29) and (30) were not made clear in the original paper (Hartley, 1948), and other quadratic estimates of ρ obtained by 'internal regression' methods have sometimes been used as alternatives to $r(1, 1)$. Finney (1958) considered the simpler equations

$$y_x = (1 - \rho) + \rho y_{x-1} \quad (35)$$

and

$$y_{x+1} - y_x = 2\alpha \frac{1 - \rho}{1 + \rho} - \frac{1 - \rho}{1 + \rho} (y_{x+1} + y_x), \quad (36)$$

which lead to $r(0, \infty)$ and $r(0, 1)$, respectively, as estimates of ρ . In a detailed investigation of the case $n = 4$ he found that Hartley's estimate of ρ showed no special advantage over $r(0, \infty)$ and $r(0, 1)$ and suggested that this result might also be true for larger values of n .

Other estimates, in the same family $r(1, k/l)$ as Hartley's estimates, can be obtained from (28) by replacing equation (29) by

$$Y_x = Y_{x-1} + ky_{x-1} + ly_x, \quad k+l \neq 0. \quad (37)$$

The estimate of ρ in this case is $(1 + kb_3)/(1 - lb_3)$, the estimate of α is $-b_2/(k+l)b_3$ and if Y_0 is taken to be zero the estimate of β is $-[b_2/(k+l)b_3] - b_1$. An estimate $r(1, k/l)$ with $k+l = 0$ exists but cannot be obtained from equation (28). In the introduction to his paper, Hartley (1948) appeared to be considering $r(1, \infty)$ rather than $r(1, 1)$ (see equation (6) of Hartley's paper). Monroe (1949), reported by Anderson (1956), developed procedures for the case $r(1, 0)$. Anderson (1956) noted that different estimates (sometimes very different) are obtained by Hartley's and Monroe's procedures. White (1956) has also considered alternative estimates in the family $r(1, k/l)$.

A more general relationship than (28) can also be used (Patterson, 1958). This is

$$y_x = b_1 + b_2\rho_0^x + b_3Y_x,$$

where

$$Y_x - \rho_0 Y_{x-1} = ky_{x-1} + ly_x.$$

The estimate of ρ in this case is $r(\rho_0, k/l)$. It is given by $(\rho_0 + kb_3)/(1 - lb_3)$. Estimates of this type have the advantage over $r(1, k/l)$ that they can be arranged to have full efficiency within the range of values of ρ occurring in practice.

NUMERICAL RESULTS

The variances and biases of the estimates $r(1, k/l)$ relative to the least-squares estimates $\hat{r}$ have been investigated numerically using the formulae quoted in previous sections and the results are set out in Tables 1-10 printed together at the end of this paper. The tabulated values S and T are defined by:

$$\text{standard error of } r = S\sigma/\beta, \quad (38)$$

$$\text{bias in } r = \mathcal{E}(r) - \rho = B\sigma^2/\beta^2 \quad (39)$$

and

$$T = B/S. \quad (40)$$

The tables are given in terms of $\theta = \rho^{n-1}$ rather than ρ because large values of ρ within the range $0 < \rho < 1$ tend to become more important as n increases. If n is changed by changing the interval but not the range of the independent variate then ρ^{n-1} remains constant. Also, $n^3 \text{ var } r$ and $n^2 \text{ bias } r$ tend to non-zero limiting values as n is increased with θ constant ($0 < \theta < 1$) both for $r = \hat{r}$ and for $r = r(1, k/l)$.

Values of S for $\hat{r}$ are set out in Table 1 for various θ and $n = 4$ to 7, 9, 12 and 20. Corresponding values for the estimates $r(1, k/l)$ are given in Table 2; these values are independent of k/l . As n increases the $S n^{\frac{3}{2}}$ tend to limiting values which are set out in Table 9. The efficiencies of $r(1, k/l)$, given by the ratios of S^2 for $\hat{r}$ and $r(1, k/l)$, are set out in Table 3 and, for very large n , in Table 4.

The estimates $r(1, k/l)$ tend towards full efficiency as θ approaches 1 but this is of no value in itself as the standard errors of even the efficient estimates tend to be very large in this

region. They are, however, also highly efficient over a wide range of smaller values of θ . In fact the efficiencies only tend to become low for large n and very small values of θ .

The generally high efficiencies of $r(1, k/l)$ compare very favourably with the efficiencies of other non-efficient estimates of ρ . Thus:

(a) The estimates $r(0, k/l)$ are very efficient when $n = 4$, but their efficiencies fall off rapidly when n is increased (Patterson, 1958).

(b) The efficiencies of linear estimates, $r(\rho_0)$, of the type considered by Patterson (1956) fall off relatively quickly as the difference between ρ and ρ_0 increases. Table 4 shows that, for very large n , $r(1, k/l)$ covers a rather wider range of θ with high efficiency than two linear estimates together: $r(1)$ and $r(\rho_0)$ with $\rho_0^{n-1} = 0.0071$. The simplicity of calculation of linear estimates is still useful in practice, however, for example as initial estimates in the iterative least-squares procedure of Stevens (1951).

(c) Estimates $r(\rho_0, k/l)$ with suitably chosen ρ_0 lying between 0 and 1 cover a wider range of θ than $r(1, k/l)$. Thus if ρ_0^{n-1} is taken to be 0.0071 the estimate is 95% efficient (for very large n) for $\theta > 0.000075$ whilst the corresponding limit for $r(1, k/l)$ is about 0.0015. Very small values of θ appear, however, to be rare in practice.

The values of T set out in Tables 5–8 give, when multiplied by σ/β , the ratios of the biases to the standard errors of $\hat{r}$ and of $r(1, k/l)$, with $k/l = 0, 1$ and ∞ . Unlike the variances the biases in $r(1, k/l)$ depend on k/l . Limiting values of $Tn^{\frac{1}{2}}$ are given in Table 9. Negative values of T indicate that ρ is underestimated.

The biases in Hartley's estimate $r(1, 1)$ (Table 7) are often considerably greater than those in the least-squares estimate $\hat{r}$ (Table 5). For many purposes, however, they can probably be ignored as they are not large unless the estimates of ρ have large standard errors and are consequently already unsatisfactory. For example, if $n = 12$, $\theta = 0.001$ (i.e. $\rho = 0.53$) and $\sigma = 0.1\beta$ the standard error of $r(1, 1)$ is ± 0.08 and the bias is about 0.01. σ/β would have to be about 0.42 for the bias to be one-half of the standard error, in which case the standard error would be as high as 0.32.

The biases in the estimates $r(1, 0)$ and $r(1, \infty)$ are generally much larger and except when $n = 4$ there is no point in using either of these estimates in preference to $r(1, 1)$. If $n = 4$ the biases in $r(1, \infty)$ are smaller than the biases in $r(1, 1)$ but even this is of little value as much simpler estimates, e.g. $r(0, 2.75)$, with as high efficiency and very low bias are available.

Bias in Hartley's estimate may become serious if averages of several independent estimates of ρ are required. The bias can, of course, be estimated from the tables, provided that an accurate estimate of σ/β is available. For this purpose σ requires to be determined from variation between replicates. This method is likely to deal satisfactorily only with small biases. If the biases are large the approximations involved in calculating the T and S may no longer be adequate and the estimate of β is likely to be poor. An alternative and generally more reliable method for dealing with bias is to replace $r(1, 1)$ by $r(1, k/l)$ with values of k/l given by Table 10. These values k/l are such that the biases in $r(1, k/l)$ are zero. Great accuracy in k/l is not required. Thus for $n > 5$ a single value of k/l , 1.5 say, can be used for most purposes.

The standard errors and biases of functions of ρ can also be obtained using the tabulated values of S and T . Thus, for example, the function $-(\ln \rho)/m$, where m is a scaling factor, is often of greater interest than ρ itself. In this case if r is an estimate of ρ and $\hat{\lambda} = -(\ln r)/m$,

$$\text{var}(\hat{\lambda}) = \left\{ \mathcal{E} \left(\frac{d\hat{\lambda}}{dr} \right) \right\}^2 \text{var}(r) = \frac{\text{var}(r)}{m^2 \rho^2}, \quad (41)$$

then

so that

$$\text{s.e.}(\hat{\lambda}) = \frac{\text{s.e.}(r)}{m\rho} = \frac{S\sigma}{m\rho\beta}. \quad (42)$$

The bias in $\hat{\lambda}$ is

$$\text{bias}(\hat{\lambda}) = \mathcal{E}(\hat{\lambda}) - \lambda = \left\{ \mathcal{E} \left(\frac{d\hat{\lambda}}{dr} \right) \right\} \text{bias}(r) + \frac{1}{2} \left\{ \mathcal{E} \left(\frac{d^2\hat{\lambda}}{dr^2} \right) \right\} \text{var}(r) \quad (43)$$

$$= -\frac{\text{bias}(r)}{m\rho} + \frac{\text{var}(r)}{2m\rho^2}, \quad (44)$$

i.e.

$$\frac{\text{bias}(\hat{\lambda})}{\text{s.e.}(\hat{\lambda})} = -\frac{\text{bias}(r)}{\text{s.e.}(r)} + \frac{\text{s.e.}(r)}{2\rho}. \quad (45)$$

In these expressions all standard errors are taken to be positive.

The efficiencies of $\hat{\lambda}$ are, of course, identical with the efficiencies of r . The loss of efficiency of Hartley's estimate for low θ may be of some importance when the estimation of λ is of primary interest. The total amount of information on λ is proportional to $\rho^2(\ln \rho)^2/\text{var } \hat{r} = I$ which has its maximum when θ is small. Thus when $n = 20$ the values of I are as follows:

θ	0.000001	0.000125	0.001	0.008	0.064
I	0.228	0.275	0.289	0.267	0.146

Hence useful estimates of λ are available when θ is as low as 10^{-6} and it would therefore be worth while in this case to use a better estimate than the 72 % efficient Hartley estimate.

As in the case of the estimates of ρ , the expectation of the Hartley estimate of λ is closer to the expectation of the least-squares estimate than are the estimates with $k/l = 0$ or ∞ . This can be readily seen from equation (45); the second term on the right-hand side is approximately the same for all four estimates.

Thus similar recommendations can be made for the estimation of both ρ and λ . If high-speed computers are available Stevens's method should be used as it ensures a good fit in all cases in which a good fit is possible. If computing facilities are limited and the ordinates are equally spaced, Hartley's method can be used and in most cases will be superior, in the sense of having high efficiency and relatively low bias, to many other inefficient estimates which have been suggested. A worthwhile modification that can be made is to replace the $k/l = 1$ in the Hartley procedure by $k/l = 1.5$ or some other convenient value determined from Table 10. This will often result in a reduction of bias. If the curvature is very marked so that most of the y 's lie near the asymptote it may be desirable to use the related estimate $r(\rho_0, k/l)$ with ρ_0 not equal to 1.

SUMMARY

The efficiencies and biases of the estimates of ρ in the equation $y = \alpha - \beta\rho^x$, obtained by the method of Hartley (1948), have been examined numerically using general formulae derived by Patterson (1958). It is shown that while Hartley's method generally leads to estimates having relatively high efficiencies and small biases, and can therefore be regarded as adequate for many purposes, the same conclusion does not hold for similar methods which are sometimes used in place of the original Hartley method.

REFERENCES

- ANDERSON, R. L. (1957). Some statistical problems in the analysis of fertilizer response data. Chapter 17 of *Fertilizer Innovations and Response Use*, Ed. E. L. Baum; Iowa State College Press.
- FINNEY, D. J. (1958). The efficiencies of alternative estimators for an asymptotic regression equation. *Biometrika*, **45**, 370-88.
- HARTLEY, H. O. (1948). The estimation of non-linear parameters by 'internal least squares'. *Biometrika*, **35**, 32-45.
- MONROE, R. J. (1949). Unpublished thesis. North Carolina State College.
- PATTERSON, H. D. (1956). A simple method for fitting an asymptotic regression curve. *Biometrics*, **12**, 323-9.
- PATTERSON, H. D. (1958). The use of autoregression in fitting an exponential curve. *Biometrika*, **45**, 389-400.
- STEVENS, W. L. (1951). Asymptotic regression. *Biometrics*, **7**, 247-67.
- WHITE, R. F. (1956). Unpublished thesis. Iowa State College.

Note added in proof

Another method of estimating ρ which deserves to be mentioned is to replace the w_x of equation (2) by suitable functions $u_x + rv_x$ and solve the resulting quadratic equation in r . In the case $n = 4$, using $u_x = -10, 5, 5$ and $v_x = -17, -5, 22$, efficiencies of at least 99.9% are obtained over the entire range of ρ and the estimates are close approximations to $\hat{r}$. For $n = 5$ suitable values of u_x and v_x are $u_x = -16, 5, 7, 4$ and $-14, -22, 10, 26$. There are grounds for believing that in general high efficiency can be obtained at least as easily by this method as by the use of quadratic estimates.

H.D.P.

Table 1. Multipliers, S , of σ/β in terms of the standard error of $\hat{r}$

θ	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 9$	$n = 12$	$n = 20$
0.0	1.225	1.155	1.118	1.095	1.069	1.049	1.028
0.000125	1.267	1.200	1.149	1.100	0.996	0.847	0.561
0.001	1.312	1.228	1.153	1.079	0.939	0.762	0.470
0.008	1.421	1.287	1.166	1.056	0.871	0.668	0.382
0.064	1.777	1.503	1.287	1.114	0.860	0.621	0.328
0.216	2.560	2.035	1.668	1.398	1.034	0.718	0.362
0.512	5.024	3.772	2.981	2.437	1.744	1.178	0.575

Table 2. Multipliers, S , of σ/β in terms of the standard error of $r(1, k/l)$

θ	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 9$	$n = 12$	$n = 20$
0.0	1.247	1.225	1.241	1.272	1.349	1.475	1.787
0.000125	1.282	1.237	1.200	1.158	1.059	0.906	0.601
0.001	1.322	1.250	1.182	1.010	0.970	0.789	0.487
0.008	1.424	1.295	1.176	1.066	0.861	0.676	0.386
0.064	1.777	1.504	1.288	1.115	0.860	0.621	0.328
0.216	2.560	2.035	1.668	1.398	1.034	0.718	0.362
0.512	5.024	3.772	2.981	2.437	1.744	1.178	0.575

Table 3. *Percentage efficiencies of $r(1, k/l)$*

θ	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 9$	$n = 12$	$n = 20$
0.0	96	89	81	74	63	51	33
0.000001	97	90	86	82	80	74	72
0.000125	98	94	92	90	88	88	87
0.001	98	96	95	94	94	93	93
0.008	99	99	98	98	98	98	98

For $\theta \geq 0.064$ the efficiency is almost 100 %.Table 4. *Percentage efficiencies of linear and quadratic estimates for very large n*

θ	Linear estimates		Quadratic estimates	
	$r(1)$	$r(\rho_0)^*$	$r(1, k/l)$	$r(\rho_0, k/l)^*$
0.000001	—	52	72	86
0.000125	—	83	88	96
0.001	—	95	94	100
0.008	82	100	98	100
0.064	94	94	99	99
0.216	98	87	100	99
0.512	99	—	100	99
1.000	100	—	100	—

* $\rho_0^{n-1} = 0.0071$.Table 5. *Multipliers T of σ/β in the ratio (bias of $\hat{r})/(\text{standard error of } \hat{r})$*

θ	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 9$	$n = 12$	$n = 20$
0.0	0.204	0.096	0.056	0.036	0.019	0.010	0.003
0.000125	0.223	0.059	-0.046	-0.127	-0.247	-0.361	-0.482
0.001	0.253	0.062	-0.061	-0.150	-0.273	-0.376	-0.464
0.008	0.344	0.119	-0.019	-0.115	-0.234	-0.323	-0.387
0.064	0.633	0.354	0.187	0.132	-0.055	-0.151	-0.225
0.216	1.134	0.743	0.511	0.359	0.175	0.037	-0.085
0.512	2.451	1.706	1.266	0.978	0.628	0.360	0.099

Table 6. Multipliers T of σ/β in the ratio $\{\text{bias of } r(1, 0)\}/\{\text{standard error of } r(1, 0)\}$

θ	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 9$	$n = 12$	$n = 20$
0.0			$(T = \infty \text{ for all } n)$		10.067	8.676	6.800
0.000125	18.283	15.141	13.039	11.700	7.138	6.360	5.144
0.001	9.479	9.051	8.417	7.184	5.052	4.742	3.942
0.008	5.249	5.650	5.608	5.472	4.156	3.903	3.316
0.064	3.585	4.122	4.268	4.278	4.436	4.187	3.579
0.216	3.777	4.315	4.493	4.527	6.741	6.370	5.456
0.512	5.975	6.626	6.846	6.885			

Table 7. Multipliers T of σ/β in the ratio $\{\text{bias of } r(1, 1)\}/\{\text{standard error of } r(1, 1)\}$

θ	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 9$	$n = 12$	$n = 20$
0.0				2.789	5.198	9.684	25.430
0.000125	0.534	1.021	1.805	1.274	1.588	1.804	1.857
0.001	0.520	0.762	1.040	0.852	1.091	1.182	1.165
0.008	0.522	0.655	0.812	0.655	0.681	0.708	0.659
0.064	0.563	0.574	0.618	0.541	0.489	0.439	0.359
0.216	0.779	0.644	0.580	0.665	0.532	0.421	0.290
0.512	1.233	0.938	0.773	1.192	0.877	0.626	0.358
	2.520	1.842	1.449				

Table 8. Multipliers T of σ/β in the ratio $\{\text{bias of } r(1, \infty)\}/\{\text{standard error of } r(1, \infty)\}$

θ	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 9$	$n = 12$	$n = 20$
0.0					0.436	2.180	9.072
0.000125	-0.356	-0.612	-0.564	-0.342	-1.169	-1.232	-1.223
0.001	-0.368	-0.759	-0.949	-1.057	-1.458	-1.581	-1.602
0.008	-0.374	-0.838	-1.098	-1.150	-1.709	-1.892	-1.888
0.064	-0.374	-0.944	-1.282	-1.500	-2.112	-2.259	-2.200
0.216	-0.343	-1.106	-1.548	-1.823	-2.692	-2.856	-2.744
0.512	-0.294	-1.364	-1.966	-2.326	-4.516	-4.778	-4.563
	-0.244	-2.204	-3.272	-3.900			

Table 9. *Values of $Sn^{\frac{3}{2}}$ and $Tn^{\frac{1}{2}}$ when n is very large*

θ	$Sn^{\frac{3}{2}}$		$Tn^{\frac{1}{2}}$			
	$\hat{r}$	$r(1, k/l)$	$\hat{r}$	$r(1, 0)$	$r(1, 1)$	$r(1, \infty)$
0.000001	159.3	187.2	-5.42	54.96	28.76	2.55
0.000125	90.2	95.1	-4.48	31.65	11.67	-8.30
0.001	66.0	68.1	-3.82	24.92	7.06	-10.79
0.008	47.0	47.4	-2.94	19.68	3.78	-12.13
0.064	35.6	35.7	-1.86	16.99	1.77	-13.46
0.216	36.8	36.8	-1.25	18.50	1.06	-16.38
0.512	55.8	55.8	-0.83	28.32	0.72	-26.91

Table 10. *Values of k/l for zero bias in $r(1, k/l)$*

θ	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 9$	$n = 12$	$n = 20$	$n = \infty$
0.0	2.50	2.67	4.20	9.14	-10.91	-3.44	-1.80	-1.00
0.000125	2.49	2.11	2.28	2.47	2.80	3.11	3.47	3.81
0.001	2.53	1.92	1.92	1.97	2.06	2.15	2.23	2.31
0.008	2.81	1.79	1.67	1.63	1.62	1.62	1.62	1.62
0.064	4.18	1.88	1.59	1.48	1.40	1.35	1.30	1.31
0.216	7.71	2.16	1.68	1.51	1.36	1.28	1.20	1.13
0.512	19.6	2.54	1.83	1.58	1.37	1.25	1.15	1.05
1.0	∞	3.00	2.00	1.67	1.40	1.25	1.12	1.00

THE EFFICIENCY OF INTERNAL REGRESSION FOR THE FITTING OF THE EXPONENTIAL REGRESSION

By H. O. HARTLEY

Statistical Laboratory, Iowa State College

Ten years ago (Hartley, 1948), I suggested a method of estimating the parameters α , β and ρ of the exponential equation

$$y = \alpha + \beta \rho^x \quad (1)$$

by what I called 'internal regression'. The notation of (1) is that used by Finney (1958) who compares my estimators with those arising from the direct least-square treatment of the auto-regression equation

$$y_{i+1} = \alpha(1 - \rho) + \rho y_i, \quad (2)$$

where the y_i are the observed values at n regularly spaced points $x = i = 1, 2, \dots, n$. Finney attributes the use of this equation to Dr St C. S. Taylor, deals fully only with the case $n = 4$, and considers both the model (a) where the errors in (1) are assumed independent and of constant variance, and (b) where this assumption is made for the errors in (2). My paper, giving examples of the response of crop-yields to independent applications of fertilizers, was mainly concerned with (a). In comparing the estimators arising from the direct least-square treatment of (2) with my method (which is essentially a least-squares method applied to the progressive sums of (2)) Finney states (p. 376): 'I have been unable to understand what particular advantage can be claimed for this procedure on the evidence that Hartley supplies. Viewed as a regression calculation, it is considerably more laborious than either of those considered above, because of the various partial sums that must be formed. So far as its precision is concerned, the construction by way of regression is irrelevant because it takes no account of the pattern of the errors.'

I do not think that the full facts of the case justify this assessment, for not only do I discuss the pattern of errors (called by me residuals, see § 3.1), but I also point out certain properties of these residuals which are indicative of the relative efficiency of various methods of estimation. It was the study of these residuals which prevented me from using the least-square estimator arising from (2) under assumption (a). That is, I deliberately decided, on the strength of the error residual relationships summarized in my paper, against using the very estimator which appears to be advocated in the above passage by Finney. This estimator has been critically examined by Patterson (1958) and is considered as severely restricted in general use by Finney himself (1958, p. 387, line 4 from bottom) after his study of Patterson's results. Moreover (p. 34, lines 6-16), I clearly state the motivation for my estimators—I am justifying their use by the evaluation of their large sample variances and efficiencies obtained by comparison with the large sample variance formulae of the maximum likelihood estimates. I agree, of course, that a good large sample performance of an estimator does not necessarily guarantee its efficiency for small samples, but I did no more than follow a common statistical practice in advocating estimation procedures on the grounds of efficient large sample performance.

For the special case $\alpha = 0$ and n large these large sample formulae are given for case (a) on pages 43-4 of my paper and graphically exhibited in Fig. 3, where it will be seen that the

efficiency of the method tends to 100 % as $\rho \rightarrow 1$ (my $q = -n \ln \rho \rightarrow 0$) and remains high for the practical range of ρ . At the time of publishing my paper similar large sample results for the estimates of all three parameters in equation (1) were in the course of being worked out by a student of mine (Miss I. Trivedi) and her results were subsequently recorded in her Master's Thesis (University College London, 1949). These results for large n again show the high efficiency of the method for the practical range of ρ .

In the present issue there appears a paper by Patterson & Lipton (1959) who kindly allowed me to have a preview of their typescript. These authors give a most comprehensive evaluation of my method of estimation for both large and small n using the elegant formulae previously derived by Patterson (1958).

It is of interest to compare the limiting efficiencies (for $n \rightarrow \infty$) obtained by them with those computed from Trivedi's approximate formulae. The results are shown in Table 1 and appear to be in good agreement.

Table 1. *Comparison of the percentage efficiency of an 'internal regression' estimator*

$q = -\ln \theta$	Computed from Trivedi's formula		Computed from Patterson's and Lipton's formula	
	θ	% efficiency	θ	% efficiency
			0.064	99
3	0.0498	99.7	—	—
4	0.01832	99.2	—	—
5	0.00674	98.0	0.008	98
6	0.002479	96.2	—	—
7	0.000912	93.8	0.001	94
8	0.000335	90.9	—	—
9	0.000123	87.8	0.000125	88
10	0.000045	84.6	—	—
			0.000001	72

However, for small values of n , Patterson's and Lipton's results on efficiency and bias could certainly not have been inferred from our earlier large sample results. Indeed, the extremely good performance of my estimator of ρ for small values of n , which was demonstrated by their calculations, was a pleasant surprise to me. White (1956) (a student of O. Kempthorne) had independently obtained some of Patterson's formulae, but had not made a comprehensive numerical evaluation of efficiencies (or bias).

Patterson and Lipton recommend the use of Stevens's (1951) 100% efficient least-squares estimator where a high speed computer is available and I agree that where computing centres are faced with the fitting of this exponential curve sufficiently frequently, the writing of a High Speed Computer Program specifically for this computation and for model (a) is warranted.

Powerful methods of analysis, such as those employed in the above paper, permit a detailed assessment of the relative merits of estimators. The question now arises, how sensitive are the more refined judgements to likely departures of the error pattern from the

assumed model. For example, when the first ordinate of the exponential curve (measured from the asymptote) is 20 times as large as the last ordinate or even larger (Patterson's & Lipton's $\theta \leq 0.05$), it is doubtful whether the assumption of constant variance will be strictly satisfied. Since it is only in this range of θ that my estimator fails to be fully efficient, a re-examination of this drop of efficiency under variance heterogeneity may be desirable. Finally, this raises the important question of the 'robustness' of the two estimators. I am sure we are all agreed that as much information on the physical mechanism (generating the responses y) as one can possibly obtain, should be used as a guide to set up an error model. Thus in independent fertilizer trials one would usually be inclined towards (a), in time series phenomena such as growth curves towards (b) or some other stochastic alternative. However, often a model (such as the model (a)) is accepted not because of definite evidence as to its validity but rather because of lack of evidence for a better alternative. In such situations we must be prepared for (at least slight) departures from the accepted model and the question of the 'robustness' of our optimum estimators becomes of great importance. Patterson (1958) seems to have made a start with such studies and reports (p. 399) that he has found Stevens's estimator fairly effective under the model (b). Finney (1958) also discusses mixtures of the two models. However, there are clearly other alternatives to the two models and much remains to be done on this issue.

REFERENCES

- FINNEY, D. J. (1958). The efficiencies of alternative estimators for an asymptotic regression equation. *Biometrika*, **45**, 370-88.
- HARTLEY, H. O. (1948). The estimation of non-linear parameters by 'internal least squares'. *Biometrika*, **35**, 32-45.
- PATTERSON, H. D. (1958). The use of autoregression in fitting an exponential curve. *Biometrika*, **45**, 389-400.
- PATTERSON, H. D. & LIPTON, S. (1959). An investigation of Hartley's method for fitting an exponential curve. *Biometrika*, **46**, 281-92.
- STEVENS, W. L. (1951). Asymptotic regression. *Biometrics*, **7**, 247-67.
- TRIVEDI, I. (1949). Unpublished Thesis, University College London.
- WHITE, R. F. (1956). Unpublished Thesis, Iowa State College.

THE DISTRIBUTION OF MOMENT ESTIMATORS

BY L. R. SHENTON

College of Science and Technology, Manchester

1. In a previous paper (Shenton, 1958) we introduced a class of moment estimator depending on the sample moments. We now consider the sampling distribution of a moment estimator and give expressions for the first four cumulants, these suggesting that the distribution is asymptotically normal. As a special case, when the moment estimators depend on an infinity of moments, the cumulants are those of the maximum likelihood estimator (assuming one exists) and our expressions agree with those given by Haldane & Smith (1956).

Our main purpose here is to treat the problem for the case of a single parameter in general and attempt an approach which will lend itself to the development of the multi-parameter estimation problem. Thus although we give an illustrative example we do this to indicate what complexity is to be expected in the method and not as a practical illustration, for most of these involve at least two parameters. The extension to several parameters involving simultaneous estimation is deferred, for there are several difficulties of complication to be overcome (see Haldane, 1953).

2.1. Let $P(x, \theta)$ be the probability of the variate x depending on the parameter θ . For the sample and population moments we write

$$m_s = \sum_{j=1}^n x_j^s / n, \quad m_s^* = \sum_{j=1}^n (x_j - m_1)^s / n, \quad (1a)$$

$$\mu_s = \mathcal{E} m_s, \quad \mu_s^* = \mathcal{E} m_s^*. \quad (1b)$$

It is assumed that the moments exist and they are differentiable in some θ -interval. It is further assumed that the range of x is independent of θ .

2.2. The q th moment estimator θ_q (when it exists) satisfies the determinantal equation

$$h(\phi) = 0, \quad \phi_i = \theta_q, \quad (2)$$

where

$$h(\theta) = \begin{vmatrix} 0 & m_0 & m_1 & \dots & m_q \\ 0 & \mu_0 & \mu_1 & \dots & \mu_q \\ \frac{\partial \mu_1}{\partial \theta} & \mu_1 & \mu_2 & \dots & \mu_{q+1} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{\partial \mu_q}{\partial \theta} & \mu_q & \mu_{q+1} & \dots & \mu_{2q} \end{vmatrix}. \quad (3)$$

This is an alternative form of the expression (5) given by Shenton (1958, p. 411) for the case of a single parameter. It may be derived by applying Schweins's theorem in determinants (see, for example, Aitken (1946) or Muir (1906)) to the series expansion (truncated) for a moment estimator appearing in (13), p. 112, of my 1950 paper. It will be noted that with the

exception of the first row in (3), $h(\theta)$ consists of elements which are functions of θ . It is convenient to expand $h(\theta)$ by its first row in terms of cofactors as follows:

$$h(\theta) = \sum_{s=0}^q m_s \mathcal{M}_s(\theta) \quad (4a)$$

$$= \langle m, \mathcal{M} \rangle. \quad (4b)$$

For brevity $\mathcal{M}_s$ is written for $\mathcal{M}_s(\theta)$, and in (4b) we have introduced a notation for the vector inner product of the vectors $m = [m_0, m_1, \dots, m_q]$ and $\mathcal{M} = [\mathcal{M}_0, \mathcal{M}_1, \dots, \mathcal{M}_q]$. It will be noted that m has q random components, for $m_0 = 1$.

2.3. From (2) we can write the stochastic q -dimensional Taylor expansion for θ_q about the point μ , and in fact

$$\theta_q = \{ \exp \langle A(\partial/\partial m) \rangle \} \theta_q |_{m=\mu, A=M}, \quad (5)$$

where $M = m - \mu$, A an arbitrary vector. After the evaluation of the inner-product operator in (5) we have

$$\theta_q = \sum_{s=0}^{\infty} \Phi_s(M, \theta), \quad (6a)$$

$$s! \Phi_s = \langle M(\bar{\partial}/\partial m) \rangle^s \theta_q, \quad (6b)$$

where

and in (6b) the partial derivatives of θ_q are to be evaluated at $m = \mu$, this being implied in the *barred* derivative symbol. For example,

$$\Phi_1 = M_1 \frac{\bar{\partial} \theta_q}{\partial m_1} + M_2 \frac{\bar{\partial} \theta_q}{\partial m_2} + \dots + M_q \frac{\bar{\partial} \theta_q}{\partial m_q},$$

$$\frac{\bar{\partial} \theta_q}{\partial m_1} = \left. \frac{\partial \theta_q}{\partial m_1} \right|_{m=\mu}, \quad \text{etc.}$$

where

2.4. *Expression for Φ_s .* Differentiating (2) partially with respect to m_s we have

$$\frac{\partial \phi}{\partial m_s} \frac{\partial h(\phi)}{\partial \phi} + \mathfrak{M}_s = 0 \quad (s = 1, 2, \dots, q; \mathfrak{M}_s = \mathcal{M}_s(\phi)), \quad (7)$$

$$\mathcal{M}_s + \bar{h}_1 \frac{\bar{\partial} \theta_q}{\partial m_s} = 0, \quad (8)$$

from which

$$\bar{h}_1 = \frac{\bar{\partial}}{\partial \phi} h(\phi) = - \left\langle \mathcal{M} \frac{\partial \mu}{\partial \theta} \right\rangle. \quad (9)$$

where from (3)

$$\Phi_1 = - \langle \mathcal{M} M \rangle / \bar{h}_1.$$

We thus have

To find an expression for Φ_2 , we require terms such as $\bar{\partial}^2 \phi / \partial m_r \partial m_s$, which could be found from (7) by further differentiation. This piecemeal procedure can be avoided, and indeed from (7), for arbitrary A

$$\frac{\partial h(\phi)}{\partial \phi} \left\langle A \frac{\partial \phi}{\partial m} \right\rangle = - \langle A \mathfrak{M} \rangle, \quad (10)$$

so that

$$\begin{aligned} \left\langle A \frac{\partial}{\partial m} \right\rangle \left\{ \frac{\partial h(\phi)}{\partial \phi} \left\langle A \frac{\partial \phi}{\partial m} \right\rangle \right\} &= - \left\langle A \frac{\partial}{\partial m} \right\rangle \langle A \mathfrak{M} \rangle \\ &= - \left\langle A \frac{\partial \mathfrak{M}}{\partial \phi} \right\rangle \left\langle A \frac{\partial \phi}{\partial m} \right\rangle, \end{aligned}$$

if A is independent of m . Thus

$$\frac{\partial \phi(\phi)}{\partial \phi} \left\langle A \frac{\partial}{\partial m} \right\rangle^2 \phi + \left\langle A \frac{\partial \phi}{\partial m} \right\rangle \left\{ \left\langle A \frac{\partial}{\partial m} \right\rangle \frac{\partial h(\phi)}{\partial \phi} \right\} = - \left\langle A \frac{\partial \mathfrak{M}}{\partial \phi} \right\rangle \left\langle A \frac{\partial \phi}{\partial m} \right\rangle. \quad (12)$$

But

$$\left\langle A \frac{\partial}{\partial m} \right\rangle \frac{\partial h(\phi)}{\partial \phi} = \left\langle A \frac{\partial \phi}{\partial m} \right\rangle \frac{\partial^2 h(\phi)}{\partial \phi^2} + \left\langle A \frac{\partial \mathfrak{M}}{\partial \phi} \right\rangle,$$

so that substituting in (12) leads to

$$\frac{\partial h(\phi)}{\partial \phi} \left\langle A \frac{\partial}{\partial m} \right\rangle^2 \phi = - \frac{\partial^2 h(\phi)}{\partial \phi^2} \left\langle A \frac{\partial \phi}{\partial m} \right\rangle^2 - 2 \left\langle A \frac{\partial \mathfrak{M}}{\partial \phi} \right\rangle \left\langle A \frac{\partial \phi}{\partial m} \right\rangle, \quad (13)$$

and on putting $A = M$ and $m = \mu$, and using (8) we have

$$2! \Phi_2 = - \frac{\bar{h}_2}{\bar{h}_1^3} \langle M \mathcal{M} \rangle^2 + \frac{2}{\bar{h}_1^2} \langle M \mathcal{M} \rangle \left\langle M \frac{\partial \mathcal{M}}{\partial \theta} \right\rangle \quad (14)$$

in which, after differentiating (3) twice

$$\bar{h}_2 = \frac{\partial^2 h(\phi)}{\partial \phi^2} = -2 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial \mu}{\partial \theta} \right\rangle - \left\langle \mathcal{M} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle. \quad (15)$$

The expressions (10) and (14) give the first two terms of the Taylor expansion for θ_q , and they involve only terms in the sample moment deviations $m_1 - \mu_1$, etc., and the population parameter θ . Expressions for Φ_3 and higher terms may be found in a similar way. Thus for Φ_3 we operate on (13) with $\langle A(\partial/\partial m) \rangle$, replace A by M and m by μ . We thus arrive at the following expressions:

$$3! \Phi_3 = \frac{\bar{h}_1 \bar{h}_3 - 3\bar{h}_2^2}{\bar{h}_1^5} \langle M \mathcal{M} \rangle^3 + \frac{9\bar{h}_2}{\bar{h}_1^4} \langle M \mathcal{M} \rangle^2 \left\langle M \frac{\partial \mathcal{M}}{\partial \theta} \right\rangle - \frac{6}{\bar{h}_1^3} \langle M \mathcal{M} \rangle \left\langle M \frac{\partial \mathcal{M}}{\partial \theta} \right\rangle^2 - \frac{3}{\bar{h}_1^3} \langle M \mathcal{M} \rangle^2 \left\langle M \frac{\partial^2 \mathcal{M}}{\partial \theta^2} \right\rangle; \quad (16a)$$

$$4! \Phi_4 = \frac{13\bar{h}_1 \bar{h}_2 \bar{h}_3 - 24\bar{h}_2^3 - \bar{h}_4 \bar{h}_1^2}{\bar{h}_1^7} \langle M \mathcal{M} \rangle^4 + \frac{96\bar{h}_2^2 - 19\bar{h}_3 \bar{h}_1}{\bar{h}_1^6} \langle M \mathcal{M} \rangle^3 \left\langle M \frac{\partial \mathcal{M}}{\partial \theta} \right\rangle + \frac{4}{\bar{h}_1^4} \langle M \mathcal{M} \rangle^3 \left\langle M \frac{\partial^3 \mathcal{M}}{\partial \theta^3} \right\rangle - \frac{117\bar{h}_2}{\bar{h}_1^5} \langle M \mathcal{M} \rangle^2 \left\langle M \frac{\partial \mathcal{M}}{\partial \theta} \right\rangle^2 - \frac{33\bar{h}_2}{\bar{h}_1^5} \langle M \mathcal{M} \rangle^3 \left\langle M \frac{\partial^2 \mathcal{M}}{\partial \theta^2} \right\rangle + \frac{42}{\bar{h}_1^4} \langle M \mathcal{M} \rangle \left\langle M \frac{\partial \mathcal{M}}{\partial \theta} \right\rangle^3 + \frac{45}{\bar{h}_1^4} \langle M \mathcal{M} \rangle^2 \left\langle M \frac{\partial \mathcal{M}}{\partial \theta} \right\rangle \left\langle M \frac{\partial^2 \mathcal{M}}{\partial \theta^2} \right\rangle; \quad (16b)$$

$$\bar{h}_3 = -3 \left\langle \frac{\partial^2 \mathcal{M}}{\partial \theta^2} \frac{\partial \mu}{\partial \theta} \right\rangle - 3 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle - \left\langle \mathcal{M} \frac{\partial^3 \mu}{\partial \theta^3} \right\rangle = \left\langle \frac{\partial^3 \mathcal{M}}{\partial \theta^3} \mu \right\rangle; \quad (16c)$$

$$\bar{h}_4 = -4 \left\langle \frac{\partial^3 \mathcal{M}}{\partial \theta^3} \frac{\partial \mu}{\partial \theta} \right\rangle - 6 \left\langle \frac{\partial^2 \mathcal{M}}{\partial \theta^2} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle - 4 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial^3 \mu}{\partial \theta^3} \right\rangle - \left\langle \mathcal{M} \frac{\partial^4 \mu}{\partial \theta^4} \right\rangle = \left\langle \frac{\partial^4 \mathcal{M}}{\partial \theta^4} \mu \right\rangle. \quad (16d)$$

It is of interest to note that the expressions in (9), (15), (16c) and (16d) follow immediately from the identity $\langle \mathcal{M} \mu \rangle = 0$.

3.1. *Expected values of combinations of the Φ 's.* We require the expected value of expressions such as $\langle AM \rangle^r \langle BM \rangle^s \dots$. Now for arbitrary A

$$\mathcal{E}\{\exp y \langle MA \rangle\} = \left\{ 1 + \frac{y^2}{2! n^2} \psi_2 + \frac{y^3}{3! n^3} \psi_3 + \dots \right\}^n, \quad (17)$$

where

$$\psi_s = \mathcal{E}\langle XA \rangle^s, \quad X_r = x^r - \mu_r \quad (r = 0, 1, \dots, q).$$

Thus

$$\left. \begin{aligned} \mathcal{E}\langle MA \rangle^2 &= \psi_2/n, \\ \mathcal{E}\langle MA \rangle^3 &= \psi_3/n, \\ \mathcal{E}\langle MA \rangle^4 &= \psi_4/n^3 + 3(n-1)\psi_2^2/n^3, \\ \mathcal{E}\langle MA \rangle^5 &= \psi_5/n^4 + 10(n-1)\psi_2\psi_3/n^4, \\ \mathcal{E}\langle MA \rangle^6 &= \psi_6/n^5 + (n-1)\{10\psi_3^2 + 15\psi_2\psi_4\}/n^5 + 15(n-1)(n-2)\psi_2^3/n^5. \end{aligned} \right\} \quad (18)$$

Moreover, it is readily seen that

$$\begin{aligned} \psi_s &= \mathcal{E}\{\Sigma A_r(x^r - \mu_r)\}^s \\ &= \langle A\mu \rangle^s - \binom{s}{1} \langle A\mu \rangle^{s-1} \langle A\mu \rangle + \binom{s}{2} \langle A\mu \rangle^{s-2} \langle A\mu \rangle^2 - \dots \\ &\quad + (-1)^{s-2} \binom{s}{s-2} \langle A\mu \rangle^2 \langle A\mu \rangle^{s-2} + (-1)^{s-1} (s-1) \langle A\mu \rangle^s, \end{aligned} \quad (19)$$

in which $\langle A\mu \rangle^r$ is a symbolic multinomial expression to be interpreted by replacing a term such as $\mu_1^l \mu_2^m$ by μ_{l+2m} . For example, when $q = 2$,

$$\langle A\mu \rangle^2 = A_0^2 + A_1^2 \mu_2 + A_2^2 \mu_4 + 2A_1 A_2 \mu_3 + 2A_0 A_2 \mu_2 + 2A_0 A_1 \mu_1.$$

3.2. *Derived expressions.* To evaluate an expression such as

$$6\mathcal{E}\langle MA \rangle \langle MB \rangle \langle MC \rangle$$

we merely express this as

$$\mathcal{E} \left\langle C \frac{\partial}{\partial A} \right\rangle \left\langle B \frac{\partial}{\partial A} \right\rangle \langle MA \rangle^3,$$

A, B, C being arbitrary vectors independent of m and θ . Thus from

$$n\mathcal{E}\langle MA \rangle^2 = \langle A\mu \rangle^2 - \langle A\mu \rangle^2 \quad (20a)$$

$$n\mathcal{E}\langle MA \rangle \langle MB \rangle = \langle A\mu \rangle \langle B\mu \rangle - \langle A\mu \rangle \langle B\mu \rangle. \quad (20b)$$

we derive

$$n\mathcal{E}\langle MA \rangle \langle MB \rangle = \langle A\mu \rangle \langle B\mu \rangle - \langle A\mu \rangle \langle B\mu \rangle. \quad (21a)$$

For example

$$n\mathcal{E}\langle \mathcal{M}M \rangle \langle BM \rangle = \langle \mathcal{M}\mu \rangle \langle B\mu \rangle, \quad (21b)$$

$$n^2\mathcal{E}\langle \mathcal{M}M \rangle \left\langle \frac{\partial \mathcal{M}}{\partial \theta} M \right\rangle = \langle \mathcal{M}\mu \rangle \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle = -\Delta \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial \mu}{\partial \theta} \right\rangle; \quad (21c)$$

$$n\mathcal{E}\langle \mathcal{M}M \rangle^2 = -\Delta \left\langle \mathcal{M} \frac{\partial \mu}{\partial \theta} \right\rangle = \Delta \bar{h}_1,$$

$$\Delta = |\mu_0, \mu_2, \dots, \mu_{2q}|.$$

where

Similarly for third-order terms we find

$$\begin{aligned} n^2\mathcal{E}\langle MA \rangle \langle MB \rangle \langle MC \rangle &= \langle A\mu \rangle \langle B\mu \rangle \langle C\mu \rangle - \langle A\mu \rangle \langle B\mu \rangle \langle C\mu \rangle \\ &\quad - \langle A\mu \rangle \langle B\mu \rangle \langle C\mu \rangle + 2\langle A\mu \rangle \langle B\mu \rangle \langle C\mu \rangle, \end{aligned} \quad (22)$$

$$n^2\mathcal{E}\langle \mathcal{M}M \rangle^2 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} M \right\rangle = \langle \mathcal{M}\mu \rangle^2 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle - \langle \mathcal{M}\mu \rangle^2 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle, \quad (23a)$$

$$n^2\mathcal{E}\langle \mathcal{M}M \rangle \left\langle \frac{\partial \mathcal{M}}{\partial \theta} M \right\rangle^2 = \langle \mathcal{M}\mu \rangle \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle^2 - 2\langle \mathcal{M}\mu \rangle \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle. \quad (23b)$$

Again, without going into details, it will be found that

$$n^3 \mathcal{E} \langle \mathcal{M} M \rangle^4 = 3(n-1) (\Delta \bar{h}_1)^2 + \langle \mathcal{M} \mu \rangle^4, \quad (24a)$$

$$n^3 \mathcal{E} \langle \mathcal{M} M \rangle^3 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} M \right\rangle = -3(n-1) \Delta^2 \bar{h}_1 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial \mu}{\partial \theta} \right\rangle + \langle \mathcal{M} \mu \rangle^3 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle - \langle \mathcal{M} \mu \rangle^3 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle. \quad (24b)$$

4.1. *Moments of θ_q .* To work out the third and fourth standardized cumulants $\gamma_1 = \kappa_3/\kappa_2^{3/2}$, $\gamma_2 = \kappa_4/\kappa_2^2$ to order $n^{-1/2}$ and n^{-1} , respectively, we require the expectations of the following terms:

$$\begin{aligned} \text{Linear} & \quad \Phi_2; \\ \text{Quadratic} & \quad \Phi_1^2, \quad \Phi_1 \Phi_2, \quad \Phi_2^2, \quad \Phi_1 \Phi_3; \\ \text{Cubic} & \quad \Phi_1^3, \quad \Phi_1^2 \Phi_2; \\ \text{Quartic} & \quad \Phi_1^4, \quad \Phi_1^3 \Phi_2, \quad \Phi_1^2 \Phi_2^2, \quad \Phi_1^3 \Phi_3. \end{aligned}$$

Omitting the algebra, we quote some of the results in Table 1.

Table 1

Symbol	Expected value
θ_q	$\theta + \frac{\Delta}{2n\bar{h}_1^2} \left\langle \mathcal{M} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle + O(n^{-2})$
Φ_1^2	$\Delta/(n\bar{h}_1)$
Φ_2^2	$\frac{\Delta \Psi}{n^2 \bar{h}_1^4} + \frac{3\Delta^2}{4n^2 \bar{h}_1^4} \left\langle \mathcal{M} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle^2 + O(n^{-3})$
$\Phi_1^2 \Phi_2$	$-\frac{3\Delta^2 \bar{h}_2}{2n^2 \bar{h}_1^3} - \frac{3\Delta^2}{n^2 \bar{h}_1^3} \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial \mu}{\partial \theta} \right\rangle + O(n^{-3})$
Φ_1^4	$\frac{3(n-1)\Delta^2}{n^3 \bar{h}_1^2} + \frac{\langle \mathcal{M} \mu \rangle^4}{n^3 \bar{h}_1^4}$
$\Phi_1^3 \Phi_2$	$-\frac{\langle \mathcal{M} \mu \rangle^3 \Delta}{n^3 \bar{h}_1^5} \left\{ 5 \left\langle \mathcal{M} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle + 6 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial \mu}{\partial \theta} \right\rangle \right\}$ $-\frac{6\Delta}{n^3 \bar{h}_1^4} \left\{ \langle \mathcal{M} \mu \rangle^2 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle - \langle \mathcal{M} \mu \rangle^2 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle \right\} + O(n^{-4})$
$\Phi_1^2 \Phi_2^2$	$\frac{\Delta^3}{n^3 \bar{h}_1^5} \left\{ \frac{15}{4} \left\langle \mathcal{M} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle^2 + \frac{3\Psi}{\Delta} \right\} + O(n^{-4})$
(In this table $\Psi = \bar{h}_1 \left\{ \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle^2 - \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle^2 \right\} - \Delta \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial \mu}{\partial \theta} \right\rangle^2$.)	

4.2. The moments

$$\mu_s^*(\theta_q) = \mathcal{E}(\theta_q - \mathcal{E}\theta_q)^s$$

are now evaluated and it appears that

$$\begin{aligned} \mu_2^*(\theta_q) &= \frac{\Delta}{n\bar{h}_1} + \frac{1}{n^2 \bar{h}_1^4} \left\{ 3\Delta \Psi + \frac{7}{2} \Delta^2 \left\langle \mathcal{M} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle^2 + 3\Delta^2 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial \mu}{\partial \theta} \right\rangle \left\langle \mathcal{M} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle \right. \\ &\quad \left. + \bar{h}_1 \Delta^2 \left\langle \mathcal{M} \frac{\partial^3 \mu}{\partial \theta^3} \right\rangle + 3\Delta^2 \bar{h}_1 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle + \bar{h}_2 \langle \mathcal{M} \mu \rangle^3 \right. \\ &\quad \left. - 2\bar{h}_1 \langle \mathcal{M} \mu \rangle^2 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle + 2\Delta \bar{h}_1^3 \right\} + O(n^{-3}), \end{aligned} \quad (25a)$$

$$\mu_3^*(\theta_q) = -\frac{\langle \mathcal{M}\mu \rangle^3}{n^2 \bar{h}_1^3} + \frac{3\Delta^2}{n^2 \bar{h}_1^3} \left\langle \mathcal{M} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle + O(n^{-3}), \quad (25b)$$

$$\begin{aligned} \mu_4^*(\theta_q) = & \frac{3(n-1)\Delta^2}{n^3 \bar{h}_1^2} + \frac{\langle \mathcal{M}\mu \rangle^4}{n^3 \bar{h}_1^4} - \frac{\Delta \langle \mathcal{M}\mu \rangle^3}{n^3 \bar{h}_1^5} \left\{ 18 \left\langle \mathcal{M} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle + 24 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial \mu}{\partial \theta} \right\rangle \right\} \\ & + \frac{45\Delta^3}{n^3 \bar{h}_1^5} \left\langle \mathcal{M} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle^2 + \frac{30\Delta^3}{n^3 \bar{h}_1^5} \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial \mu}{\partial \theta} \right\rangle \left\langle \mathcal{M} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle + \frac{30\Delta^2 \Psi}{n^3 \bar{h}_1^5} \\ & - \frac{24\Delta}{n^3 \bar{h}_1^4} \left\{ \langle \mathcal{M}\mu \rangle^2 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle - \langle \mathcal{M}\mu \rangle^2 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle \right\} \\ & + \frac{10\Delta^3}{n^3 \bar{h}_1^5} \left\{ 3 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle + \left\langle \mathcal{M} \frac{\partial^3 \mu}{\partial \theta^3} \right\rangle \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle \right\} + O(n^{-4}). \quad (25c) \end{aligned}$$

4.3. We now find as first approximations to the standardized cumulants

$$\gamma_1(\theta_q) \sim \frac{3\Delta^2 \left\langle \mathcal{M} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle - \langle \mathcal{M}\mu \rangle^3}{\sqrt{(n\Delta^3 \bar{h}_1^3)}}, \quad (26a)$$

$$\begin{aligned} n\gamma_2(\theta_q) \sim & \frac{\langle \mathcal{M}\mu \rangle^4}{\Delta^2 \bar{h}_1^2} - \frac{12\langle \mathcal{M}\mu \rangle^3}{\Delta \bar{h}_1^3} \left\{ \left\langle \mathcal{M} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle + \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial \mu}{\partial \theta} \right\rangle \right\} + \frac{12\Psi}{\bar{h}_1^3} \\ & + \frac{24\Delta}{\bar{h}_1^3} \left\langle \mathcal{M} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle^2 + \frac{12\Delta}{\bar{h}_1^3} \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial \mu}{\partial \theta} \right\rangle \left\langle \mathcal{M} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle + \frac{4\Delta}{\bar{h}_1^2} \left\langle \mathcal{M} \frac{\partial^3 \mu}{\partial \theta^3} \right\rangle \\ & + \frac{12\Delta}{\bar{h}_1^2} \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle - \frac{12}{\Delta \bar{h}_1^2} \langle \mathcal{M}\mu \rangle^2 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle + 9, \quad (26b) \end{aligned}$$

with a clear indication of asymptotic normality. If $P(x, \theta)$ is linear in θ (a case which would arise for a Gram-Charlier distribution, the parameter governing the *corrective term*) there is a considerable simplification in the moments, and indeed

$$\mu_1(\theta_q) = \theta + O(n^{-2}), \quad (27a)$$

$$\mu_2^*(\theta_q) \sim \frac{\Delta}{n\bar{h}_1} + \frac{1}{n^2 \bar{h}_1^4} \left\{ 3\Delta\Psi - 2 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial \mu}{\partial \theta} \right\rangle \langle \mathcal{M}\mu \rangle^3 - 2\bar{h}_1 \langle \mathcal{M}\mu \rangle^2 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle + 2\Delta \bar{h}_1^3 \right\}, \quad (27b)$$

$$\mu_2^*(\theta_q) \sim \frac{\Delta}{n\bar{h}_1} + \frac{1}{n^2 \bar{h}_1^4} \left\{ 3\Delta\Psi - 2 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial \mu}{\partial \theta} \right\rangle \langle \mathcal{M}\mu \rangle^3 - 2\bar{h}_1 \langle \mathcal{M}\mu \rangle^2 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle + 2\Delta \bar{h}_1^3 \right\}, \quad (27c)$$

$$\gamma_1(\theta_q) \sim \frac{-\langle \mathcal{M}\mu \rangle^3}{\sqrt{(n\Delta^3 \bar{h}_1^3)}},$$

$$n\gamma_2(\theta_q) \sim \frac{\langle \mathcal{M}\mu \rangle^4}{\Delta^2 \bar{h}_1^2} - \frac{12\langle \mathcal{M}\mu \rangle^3 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial \mu}{\partial \theta} \right\rangle}{\Delta \bar{h}_1^3} + \frac{12\Psi}{\bar{h}_1^3} - \frac{12\langle \mathcal{M}\mu \rangle^2 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle}{\Delta \bar{h}_1^2} + 9. \quad (27d)$$

It is of some interest to observe that the first-order term in $\mathcal{E}(\theta_q)$ vanishes when the probability is linear in θ ; according to Haldane & Smith (1956) a similar property holds for maximum likelihood estimators in the case of linearity. We should expect this to be the case since (27a) holds for moment estimators of all orders (i.e. $q = 1, 2, \dots$) under certain regularity conditions. Haldane has suggested that estimators with this property should be described as being *almost unbiased*.

5. *Maximum likelihood estimators.* If there is a moment estimator which gives the maximum likelihood solution (this may be the case for finite q , and in any event for $q \rightarrow \infty$)

under certain limitations), then the expressions in (25)–(27) can be given in terms of the expected values of derivatives of $\log P(x, \theta)$. Thus (using $\doteq$ to indicate correspondence)

$$-\left\langle \mathcal{M} \frac{\partial \mu}{\partial \theta} \right\rangle / \Delta \doteq \mathcal{E} \frac{1}{P^2} \left(\frac{\partial P}{\partial \theta} \right)^2, \quad (28a)$$

$$-\left\langle \mathcal{M} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle / \Delta \doteq \mathcal{E} \frac{1}{P^2} \frac{\partial P}{\partial \theta} \frac{\partial^2 P}{\partial \theta^2}, \quad (28b)$$

$$\left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle^2 \doteq \mathcal{E} \left(\frac{\partial \Delta}{\partial \theta} \frac{1}{P} \frac{\partial P}{\partial \theta} - \frac{\Delta}{P^2} \left(\frac{\partial P}{\partial \theta} \right)^2 + \frac{\Delta}{P} \frac{\partial^2 P}{\partial \theta^2} \right)^2, \quad (29)$$

$$\left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial \mu}{\partial \theta} \right\rangle \doteq -\frac{\partial \Delta}{\partial \theta} \mathcal{E} \left(\frac{\partial \log P}{\partial \theta} \right)^2 + \Delta \mathcal{E} \left(\frac{\partial \log P}{\partial \theta} \right)^3 - \Delta \mathcal{E} \frac{1}{P^2} \frac{\partial P}{\partial \theta} \frac{\partial^2 P}{\partial \theta^2}, \quad (30)$$

$$\left\langle \frac{\partial \mathcal{M}}{\partial \theta} \frac{\partial^2 \mu}{\partial \theta^2} \right\rangle \doteq -\frac{\partial \Delta}{\partial \theta} \mathcal{E} \frac{1}{P^2} \frac{\partial P}{\partial \theta} \frac{\partial^2 P}{\partial \theta^2} + \Delta \mathcal{E} \frac{1}{P^3} \left(\frac{\partial P}{\partial \theta} \right)^2 \frac{\partial^2 P}{\partial \theta^2} - \Delta \mathcal{E} \frac{1}{P^2} \left(\frac{\partial^2 P}{\partial \theta^2} \right)^2, \quad (31)$$

$$\langle \mathcal{M} \mu \rangle^2 \left\langle \frac{\partial \mathcal{M}}{\partial \theta} \mu \right\rangle \doteq \mathcal{E} \left\{ -\Delta^2 \frac{\partial \Delta}{\partial \theta} \left(\frac{\partial \log P}{\partial \theta} \right)^3 + \Delta^3 \left(\frac{\partial \log P}{\partial \theta} \right)^4 - \frac{\Delta^3}{P^3} \left(\frac{\partial P}{\partial \theta} \right)^2 \frac{\partial^2 P}{\partial \theta^2} \right\}. \quad (32)$$

The derivation of most of these is straightforward, but for (28) we note that

$$\frac{\partial \log P}{\partial \theta} \doteq -\langle \mathcal{M} \mathbf{x} \rangle / \Delta,$$

where

$$\langle \mathcal{M} \mathbf{x} \rangle = \mathcal{M}_0 + \mathcal{M}_1 x + \mathcal{M}_2 x^2 + \dots,$$

and (29) is derived by differentiating (27a), and (30) from (27b). For the sampling moments of the maximum likelihood estimator $\hat{\theta}$ we therefore find (after some non-trivial algebra)

$$\mathcal{E}(\hat{\theta}) = \theta - \frac{\mathcal{E} \left(\frac{1}{P^2} \frac{\partial P}{\partial \theta} \frac{\partial^2 P}{\partial \theta^2} \right)}{2nI^2} + O(n^{-2}), \quad (33a)$$

$$\begin{aligned} \sigma^2(\hat{\theta}) = \frac{n-1}{n^2 I} + \mathcal{E} \left\{ \left(\frac{\partial \log P}{\partial \theta} \right)^4 - \frac{1}{P^3} \left(\frac{\partial P}{\partial \theta} \right)^2 \frac{\partial^2 P}{\partial \theta^2} - \frac{1}{P^2} \frac{\partial P}{\partial \theta} \frac{\partial^3 P}{\partial \theta^3} \right\} / (n^2 I^3) \\ + \left\{ \frac{7}{2} \left(\mathcal{E} \frac{1}{P^2} \frac{\partial P}{\partial \theta} \frac{\partial^2 P}{\partial \theta^2} \right)^2 - \left\{ \mathcal{E} \left(\frac{\partial \log P}{\partial \theta} \right)^3 \right\}^2 \right\} / (n^2 I^4) + O(n^{-3}), \end{aligned} \quad (33b)$$

$$\gamma_1(\hat{\theta}) = \left\{ \mathcal{E} \left(\frac{\partial \log P}{\partial \theta} \right)^3 - 3 \mathcal{E} \left(\frac{1}{P^2} \frac{\partial P}{\partial \theta} \frac{\partial^2 P}{\partial \theta^2} \right) \right\} / \sqrt{(n I^3)} + O(n^{-\frac{1}{2}}), \quad (33c)$$

$$\begin{aligned} n\gamma_2(\hat{\theta}) = \mathcal{E} \left(\frac{\partial \log P}{\partial \theta} \right)^4 / I^2 + \left\{ 24 \left(\mathcal{E} \frac{1}{P^2} \frac{\partial P}{\partial \theta} \frac{\partial^2 P}{\partial \theta^2} \right)^2 - 12 \mathcal{E} \left(\frac{\partial \log P}{\partial \theta} \right)^3 \mathcal{E} \left(\frac{1}{P^2} \frac{\partial P}{\partial \theta} \frac{\partial^2 P}{\partial \theta^2} \right) \right. \\ \left. - 4 \mathcal{E} \left(\frac{1}{P^2} \frac{\partial P}{\partial \theta} \frac{\partial^3 P}{\partial \theta^3} \right) \right\} / I^3 - 3 + O(n^{-1}). \end{aligned} \quad (33d)$$

where

$$I = \mathcal{E} \left(\frac{\partial \log P}{\partial \theta} \right)^2.$$

These expressions agree exactly (allowing for notational differences) with those given in (6) of Haldane & Smith (1956, p. 100). It is to be remarked that the forms given are not necessarily the simplest, and that there are alternatives for such terms as $\mathcal{E}\{(1/P^2) (\partial P/\partial \theta) (\partial^2 P/\partial \theta^2)\}$ in terms of the expected value of powers of $(\partial \log P/\partial \theta)$ or their derivatives (in this connexion reference may be made to Bartlett (1953) who gives an ingenious method of deriving such identities).

6. *Illustrative example.* In Shenton (1958) the asymptotic variances are given for the first three moment estimators of θ in

$$P_x = e^{-x}\{2 - \theta + (\theta - 1)x\} \quad (0 \leq x < \infty, 1 \leq \theta < 2).$$

Using the results in (27) and (33) we find for γ_1 and γ_2 for the first three moment estimators and the maximum likelihood estimator the following:

$$\left. \begin{aligned} \sqrt{n} \gamma_1(\theta_1) &\sim (2\theta^3 - 12\theta^2 + 24\theta - 12)/(-\theta^2 + 4\theta - 2)^{\frac{3}{2}}, \\ n\gamma_2(\theta_1) &\sim (-6\theta^4 + 48\theta^3 - 144\theta^2 + 192\theta - 84)/(-\theta^2 + 4\theta - 2)^2; \end{aligned} \right\} \quad (34a)$$

$$\left. \begin{aligned} \sqrt{n} \gamma_1(\theta_2) &\sim (128\theta^6 - 864\theta^5 + 2184\theta^4 - 2606\theta^3 + 1476\theta^2 - 300\theta - 16)/(\alpha\beta)^{\frac{3}{2}}, \\ n\gamma_2(\theta_2) &\sim (-6144\theta^9 + 59,904\theta^8 - 243,456\theta^7 + 568,896\theta^6 - 901,944\theta^5 + 1,064,706\theta^4 \\ &\quad - 934,512\theta^3 + 558,144\theta^2 - 194,928\theta + 29,340)/(\alpha^3\beta^2), \end{aligned} \right\} \quad (34b)$$

$$\text{where} \quad \alpha = 4\theta - 3, \quad \beta = -4\theta^3 + 15\theta^2 - 12\theta + 2; \quad (34c)$$

$$\left. \begin{aligned} \sqrt{n} \gamma_1(\theta_3) &\sim p/(qr)^{\frac{3}{2}}, \\ n\gamma_2(\theta_3) &\sim p'/(q^3r^2), \end{aligned} \right\}$$

$$\text{where} \quad p = 6750\theta^9 - 51,300\theta^8 + 180,504\theta^7 - 411,280\theta^6 + 689,472\theta^5$$

$$\quad - 854,400\theta^4 + 738,688\theta^3 - 411,840\theta^2 + 131,712\theta - 18,304,$$

$$q = 15\theta^2 - 20\theta + 6, \quad r = -15\theta^4 + 56\theta^3 - 48\theta^2 + 8,$$

$$p' = -4,556,250\theta^{14} + 52,245,000\theta^{13} - 284,318,100\theta^{12} + 1,473,232,320\theta^{11}$$

$$\quad - 7,531,774,380\theta^{10} + 28,692,814,224\theta^9 - 75,171,819,960\theta^8 + 137,243,452,800\theta^7$$

$$\quad - 178,312,654,848\theta^6 + 166,526,475,264\theta^5 - 111,281,505,024\theta^4 + 52,015,042,560\theta^3$$

$$\quad - 16,167,313,152\theta^2 + 3,002,741,760\theta - 252,062,208;$$

$$\left. \begin{aligned} \sqrt{n} \gamma_1(\theta) &\sim (\alpha + 1)\{(\alpha + 4)I - (\alpha + 1)^2/\alpha\}/I^{\frac{3}{2}}, \\ &\sim -(\alpha \log(1/\alpha))^{-\frac{3}{2}} \quad (\theta = 2 \text{ approx.}), \\ n\gamma_2(\theta) &\sim -3 + (\alpha + 1)^2\{(\alpha + 3)(\alpha + 7)I - (\alpha + 1)^2(\alpha^2 + 6\alpha - 1)/\alpha^2\}/(2I^2), \\ &\sim (2\alpha^2 \log(1/\alpha))^{-2} \quad (\theta = 2 \text{ approx.}), \end{aligned} \right\} \quad (34d)$$

$$\text{where} \quad \alpha = \frac{2 - \theta}{\theta - 1}, \quad I = -(\alpha + 1)^2 - (\alpha + 1)^3 e^\alpha Ei(-\alpha), \quad Ei(-x) = \int_{\infty}^x e^{-u} \frac{du}{u}.$$

It will be fairly evident from (34) that an investigation of the sampling distributions of moment estimators such as θ_4 (or higher orders) would be a considerable undertaking. An impression of the situation for this example is given in Figs. 1-3.* The gain in using higher order estimators is evident from a glance at the variances in Fig. 1. As for γ_1 and γ_2 , the approach of the second and third estimators to the maximum likelihood estimator is illustrated in Figs. 2 and 3. The critical value of the parameter is $\theta = 2$, for in this case the integrals appearing in the moments of the maximum likelihood estimator diverge.

7. *Conclusion.* Expressions have been given for the first two standardized cumulants γ_1 and γ_2 of the sampling distribution of the q th moment estimator. As a special case the maximum likelihood estimator is included, the results found agreeing with those of Haldane & Smith (1956).

* I am indebted to Mr A. Fletcher for assisting in the construction of these diagrams.

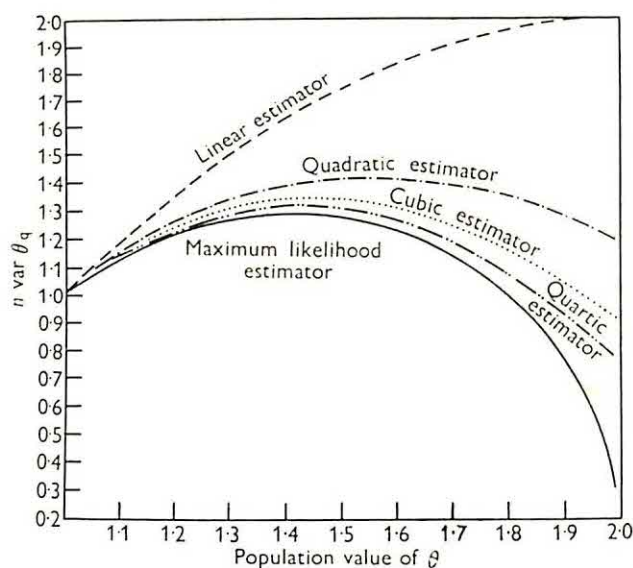


Fig. 1. Asymptotic variance of estimators of θ in $P_x = e^{-x}\{2 - \theta + (\theta - 1)x\}$.

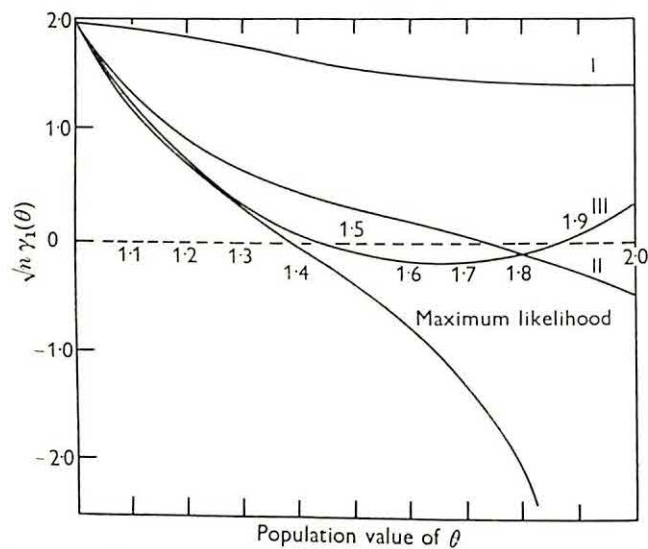


Fig. 2. $\sqrt{n} \gamma_1(\theta)$ for several estimators in $P_x = e^{-x}\{2 - \theta + (\theta - 1)x\}$.

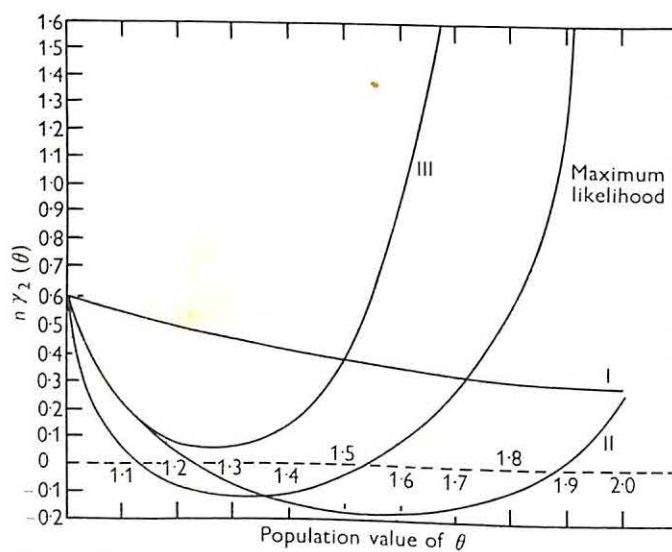


Fig. 3. $n \gamma_2(\theta)$ for several estimators in $P_x = e^{-x}\{2 - \theta + (\theta - 1)x\}$.

REFERENCES

- AITKEN, A. C. (1946). *Determinants and Matrices*. Edinburgh: Oliver and Boyd.
- BARTLETT, M. S. (1953). Approximate confidence intervals. *Biometrika*, **40**, 12-19.
- HALDANE, J. B. S. (1953). The estimation of two parameters from a sample. *Sankhya*, **12**, 313-20.
- HALDANE, J. B. S. & SMITH, SHEILA MAYNARD (1956). The sampling distribution of a maximum-likelihood estimate. *Biometrika*, **43**, 96-103.
- MUIR, T. (1906). *The Theory of Determinants*. London: Macmillan.
- SHENTON, L. R. (1950). Maximum likelihood and the efficiency of the method of moments. *Biometrika*, **37**, 111-16.
- SHENTON, L. R. (1958). Moment estimators and maximum likelihood. *Biometrika*, **45**, 311-20.

EFFICIENT ESTIMATION OF PARAMETERS IN MOVING-AVERAGE MODELS

BY J. DURBIN

Research Techniques Division, London School of Economics

1. INTRODUCTION

Although the moving-average process is one of the basic models of time-series analysis, efforts at efficient estimation of its parameters have not been very successful owing to the intractability of the maximum likelihood equations. In this paper a simple estimation procedure is suggested which in large samples gives estimates whose efficiency is as close to unity as is desired. The limiting variance matrix of the estimates is evaluated explicitly in terms of the coefficients of the model. The idea underlying the method is to fit a high-order autoregressive scheme and to base the estimates of the moving-average parameters on the fitted autoregression coefficients. Reference may be made to Wold (1938) and Whittle (1951, 1953) for earlier work on the problem.

2. MAXIMUM LIKELIHOOD APPLIED TO THE FIRST-ORDER MODEL

Suppose we have a stationary first-order moving-average model

$$x_t = \epsilon_t + \beta \epsilon_{t-1} \quad (t = 1, \dots, n), \quad (1)$$

where $\{\epsilon_t\}$ is a series of independent normal variates with zero mean and variance σ^2 , and where $|\beta| < 1$. The variance matrix of $x_1, \dots, x_n$ is $\sigma^2 V_n$ where

$$V_n = \begin{bmatrix} 1 + \beta^2 & \beta & 0 & \dots & 0 \\ \beta & 1 + \beta^2 & \beta & & \vdots \\ 0 & \beta & 1 + \beta^2 & & \vdots \\ \vdots & & & \ddots & \beta \\ 0 & \dots & \dots & \beta & 1 + \beta^2 \end{bmatrix}.$$

Using a result due to Dixon (1944) the determinant of this is $|V_n| = (1 - \beta^{2n+2})/(1 - \beta^2)$ which tends to $1/(1 - \beta^2)$ for large n . The inverse of V_n is approximately

$$\frac{1}{1 - \beta^2} \begin{bmatrix} 1 & -\beta & \beta^2 & \dots & (-\beta)^{n-1} \\ -\beta & 1 & -\beta & & \vdots \\ \beta^2 & -\beta & 1 & & \vdots \\ \vdots & & & \ddots & -\beta \\ (-\beta)^{n-1} & \dots & \dots & -\beta & 1 \end{bmatrix}.$$

Thus the likelihood is approximately given by

$$\log L = \text{constant} - \frac{1}{2} \log(1 - \beta^2) - \frac{1}{2\sigma^2(1 - \beta^2)} [\sum x_t^2 - 2\beta \sum x_t x_{t+1} + 2\beta^2 \sum x_t x_{t+2} - \dots].$$

Since the term $\frac{1}{2} \log(1 - \beta^2)$ is of small order in n compared with the remaining part of $\log L$ we may neglect it to give the approximate maximum likelihood equation

$$\frac{\partial}{\partial \beta} \left[\frac{1}{1 - \beta^2} \{ \sum x_t^2 - 2\beta \sum x_t x_{t+1} + 2\beta^2 \sum x_t x_{t+2} - \dots \} \right] = 0 \quad (2)$$

(cf. Whittle, 1951, equation 7.523). On attempting to perform the differentiation, however, we find ourselves with an unmanageable estimating equation.

On the other hand, the simple estimate obtained by equating the theoretical value of the first autocorrelation coefficient $\rho_1 = \beta/(1 + \beta^2)$ to its empirical counterpart

$$r_1 = n \sum x_t x_{t+1} / \{(n-1) \sum x_t^2\}$$

is known to be extremely inefficient. For example, for $\beta = \frac{1}{2}$ Whittle (1953) calculated its asymptotic variance to be 3.8 times that of the maximum likelihood estimate. Whittle accordingly suggests an adjustment procedure intended to bring the value closer to the maximum likelihood value but this appears to be rather complicated.

In the next section a simple but efficient alternative method is suggested.

3. ESTIMATION BASED ON THE AUTOREGRESSIVE REPRESENTATION

It is well known (see, for example, Wold, 1938) that the model (1) has the infinite autoregressive representation

$$x_t + \alpha'_1 x_{t-1} + \alpha'_2 x_{t-2} + \dots = \epsilon_t, \quad (3)$$

where $\alpha'_i = (-\beta)^i$. The remainder after $k+1$ terms of the series $x_t + \alpha'_1 x_{t-1} + \dots$ is

$$(-\beta)^{k+1} (x_{t-k-1} - \beta x_{t-k-2} + \dots) = (-\beta)^{k+1} \epsilon_{t-k-1},$$

which has variance $\beta^{2k+2} \sigma^2$. This $\rightarrow 0$ rapidly as $k \rightarrow \infty$ since $|\beta| < 1$. Consequently the finite representation

$$x_t + \alpha_1 x_{t-1} + \dots + \alpha_k x_{t-k} = \epsilon_t$$

can be made as accurate as we please by taking k sufficiently large. However, it is important to stress that although k is taken to be large, we shall always in asymptotic arguments regard it as small compared with n .

Let $a_1, \dots, a_k$ be the least-squares estimators of $\alpha_1, \dots, \alpha_k$, i.e. the estimators obtained by minimizing $\sum_{t=k+1}^n (x_t - \alpha_1 x_{t-1} - \dots - \alpha_k x_{t-k})^2$. From the results of Mann & Wald (1943) we know that $a_1, \dots, a_k$ are asymptotically normal with means $\alpha_1, \dots, \alpha_k$ and variance matrix V_k^{-1}/n , where $\sigma^2 V_k$ is the variance matrix of $x_{t-1}, \dots, x_{t-k}$. Consequently, $a_1, \dots, a_k$ have the asymptotic distribution

$$dP = \frac{n^{\frac{1}{2}} |V_k|^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}k}} \exp \left[-\frac{n}{2} \left\{ (1 + \beta^2) \sum_{i=1}^k (a_i - \alpha_i)^2 + 2\beta \sum_{i=1}^{k-1} (a_i - \alpha_i)(a_{i+1} - \alpha_{i+1}) \right\} \right] da_1 \dots da_k. \quad (5)$$

Since $\alpha_1, \dots, \alpha_k$ are autoregression coefficients they satisfy the relations

$$\begin{aligned} \alpha_1 c_0 + \alpha_2 c_1 + \dots + \alpha_k c_{k-1} &= -c_1, \\ \alpha_1 c_1 + \alpha_2 c_0 + \dots + \alpha_k c_{k-2} &= -c_2, \\ &\vdots \\ \alpha_1 c_{k-1} + \dots + \alpha_k c_0 &= -c_k, \end{aligned}$$

where $\sigma^2 c_r = E(x_t x_{t+r})$. Putting $c_0 = (1 + \beta^2) \sigma^2$, $c_1 = \beta \sigma^2$ and $c_r = 0$ ($r > 1$), we obtain

$$\begin{aligned} (1 + \beta^2) \alpha_1 + \beta \alpha_2 &= -\beta, \\ \beta \alpha_{r-1} + (1 + \beta^2) \alpha_r + \beta \alpha_{r+1} &= 0 \quad (r = 2, \dots, k-1), \\ \beta \alpha_{k-1} + (1 + \beta^2) \alpha_k &= 0. \end{aligned}$$

Multiplying these equations by $-2a_i + \alpha_i$ ($i = 1, \dots, k$) in turn and adding, we get for the quadratic expression, Q say, in the exponent of (5),

$$\begin{aligned} Q &= (1 + \beta^2) \sum_{i=1}^k (a_i - \alpha_i)^2 + 2\beta \sum_{i=1}^{k-1} (a_i - \alpha_i)(a_{i+1} - \alpha_{i+1}) \\ &= (1 + \beta^2) \sum_{i=1}^k a_i^2 + 2\beta \sum_{i=1}^{k-1} a_i a_{i+1} + 2\beta a_1 - \beta \alpha_1. \end{aligned}$$

Since, for large k , α_1 is nearly equal to $-\beta$ this gives, on putting $a_0 = 1$,

$$Q = (1 + \beta^2) \sum_{i=0}^k a_i^2 + 2\beta \sum_{i=0}^{k-1} a_i a_{i+1} - 1, \quad (6)$$

to a high degree of accuracy.

We proceed to estimate β by maximizing the likelihood obtained from the distribution of $a_1, \dots, a_k$. As in § 2, $|V_k| = (1 - \beta^{2k+2})/(1 - \beta^2)$ which for sufficiently large k is approximately equal to $1/(1 - \beta^2)$. This, however, is $O(1)$ whereas the exponent of (5) is $O(n)$. Consequently, to a first approximation maximizing the likelihood is equivalent to minimizing the quadratic form Q . Differentiating Q with respect to β and equating to zero we have for the estimator of β ,

$$b = - \frac{\sum_{i=0}^{k-1} a_i a_{i+1}}{\sum_{i=0}^k a_i^2}. \quad (7)$$

This estimator is manifestly much simpler to work with than the maximum likelihood estimator obtained from (2). As n increases it converges in probability to $-\sum_{i=0}^{k-1} \alpha_i \alpha_{i+1} / \sum_{i=0}^k \alpha_i^2$, where $\alpha_0 = 1$, and this may be made as close to β as we please by taking k sufficiently large. Since there is one term more in the denominator of (7) than in the numerator there is, perhaps, something to be said for dropping one of these terms, i.e. taking $b^* = -\sum_{i=0}^{k-1} a_i a_{i+1} / \sum_{i=0}^{k-1} a_i^2$ as the estimator. This possibility will not, however, be pursued here. A further refinement would be to take account of the determinant $|V_k|$ in (5). This would give the estimating equation

$$\frac{\beta}{n(1 - \beta^2)} - (\beta \sum a_i^2 + \sum a_i a_{i+1}) = 0$$

which can be solved iteratively. However, the effect of the extra term is generally negligible unless β^2 is close to one.

4. EFFICIENCY OF THE ESTIMATOR

The minimum asymptotic variance of consistent estimators of β has been evaluated by Whittle (1953) as $(1 - \beta^2)/n$. Whittle also shows that this minimum is attained by the maximum likelihood estimator derived from (2). We now demonstrate that the asymptotic variance of b can be made as close to the minimum as we please by taking k sufficiently large. This will be done by an argument similar to that usually employed for demonstrating the efficiency of maximum likelihood estimators (see, e.g. Cramér, 1946, § 33.3).

For large k we may take the asymptotic distribution of $a_1, \dots, a_k$ to be

$$dP = \frac{n^{\frac{1}{2}}(1 - \beta^2)^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}k}} \exp(-\frac{1}{2}nQ) da_1 \dots da_k, \quad (8)$$

where Q is given by (6). Let us write this in the form

$$dP = (1 - \beta^2)^{-\frac{1}{2}} f(Q) da,$$

and let us write the integral of this with respect to $a_1, \dots, a_k$ as

$$\begin{aligned} \int (1 - \beta^2)^{-\frac{1}{2}} f(Q) da &= 1, \\ \int f(Q) da &= (1 - \beta^2)^{\frac{1}{2}}. \end{aligned} \quad (9)$$

i.e.

Differentiating (9) and dividing through by n we have

$$\int \frac{\partial Q}{\partial \beta} f(Q) da = O\left(\frac{1}{n}\right), \quad \text{i.e.} \quad E\left[\frac{\partial Q}{\partial \beta}\right] = O\left(\frac{1}{n}\right), \quad (10)$$

and differentiating again.

$$\int \left\{ \frac{\partial^2 Q}{\partial \beta^2} - \frac{n}{2} \left(\frac{\partial Q}{\partial \beta} \right)^2 \right\} f(Q) da = O\left(\frac{1}{n}\right), \quad \text{i.e.} \quad E\left(\frac{\partial^2 Q}{\partial \beta^2}\right) = \frac{n}{2} E\left(\frac{\partial Q}{\partial \beta}\right)^2 + O\left(\frac{1}{n}\right). \quad (11)$$

Now

$$\frac{\partial Q}{\partial \beta} = 2\beta \sum_{i=0}^k a_i^2 + 2 \sum_{i=0}^{k-1} a_i a_{i+1} \quad \text{and} \quad \frac{\partial^2 Q}{\partial \beta^2} = 2 \sum_{i=0}^k a_i^2.$$

Using these results together with (7), we find

$$b - \beta = - \frac{\partial Q / \partial \beta}{\partial^2 Q / \partial \beta^2},$$

$$V(b) = \frac{E(\partial Q / \partial \beta)^2}{E(\partial^2 Q / \partial \beta^2)^2}$$

so that

to the first order in n . Observing that $E(\partial^2 Q / \partial \beta^2)^2 = [E(\partial^2 Q / \partial \beta^2)]^2$ to the first order, and using (11) we have the asymptotic result

$$V(b) = \frac{2}{nE(\partial^2 Q / \partial \beta^2)}.$$

To the first order in n , $E\left(\frac{\partial^2 Q}{\partial \beta^2}\right) = 2 \sum_{i=0}^k \alpha_i^2$, which for large k tends to $2 \sum_{i=0}^{\infty} (-\beta)^{2i} = 2/(1 - \beta^2)$.

Thus for sufficiently large k the asymptotic variance of b is

$$V(b) = \frac{1 - \beta^2}{n} \quad (12)$$

as closely as we please.

5. HIGHER-ORDER PROCESSES

The extension to higher-order processes follows along similar lines. Suppose the model is

$$x_t = \epsilon_t + \beta_1 \epsilon_{t-1} + \dots + \beta_h \epsilon_{t-h}, \quad (13)$$

where the ϵ 's are as before and where the roots of the equation $x^h + \beta_1 x^{h-1} + \dots + \beta_h = 0$ each have modulus less than one. (13) can be approximated to any required degree of accuracy by the finite autoregressive process

$$x_t + \alpha_1 x_{t-1} + \dots + \alpha_k x_{t-k} = \epsilon_t.$$

As before the asymptotic distribution of the least-squares estimates $a_1, \dots, a_k$ of $\alpha_1, \dots, \alpha_k$ is, approximately for large k ,

$$dP = \frac{n^{\frac{1}{2}} |B|^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}k}} \exp(-\frac{1}{2}nQ) da_1 \dots da_k,$$

where, writing a, α for the vectors $\{a_1, \dots, a_k\}, \{\alpha_1, \dots, \alpha_k\}$ and $\sigma^2 B$ for the variance matrix of $x_{t-1}, \dots, x_{t-k}$,

$$\begin{aligned} Q &= (a - \alpha)' B (a - \alpha) \\ &= a' B a - 2a' B \alpha + \alpha' B \alpha. \end{aligned}$$

Since α is a vector of autoregression coefficients it satisfies the equation $B\alpha + c = 0$, where $c = \{c_1, \dots, c_k\}$ and $\sigma^2 c_i = E(x_t x_{t+i})$. Thus

$$Q = a' B a + 2a' c - \alpha' c,$$

where $\alpha' c + 1 + \sum_{i=1}^h \beta_i^2$ is the constant term in the expansion of

$$(1 + \alpha_1 z + \dots + \alpha_k z^k) \sum_{i=-h}^h c_i z^i.$$

Now

$$\sum_{-h}^h c_i z^i = (1 + \beta_1 z + \dots + \beta_h z^h) (1 + \beta_1 z^{-1} + \dots + \beta_h z^{-h})$$

and

$$1 + \alpha_1 z + \dots + \alpha_k z^k = (1 + \beta_1 z + \dots + \beta_h z^h)^{-1}$$

to any required degree of accuracy by taking k sufficiently large. Thus $\alpha' c + 1 + \sum_{i=1}^h \beta_i^2 = 1$, so that

$$Q = a' B a + 2a' c + \sum_{i=1}^h \beta_i^2. \quad (14)$$

As for the first-order process $\log |B|$ is found to be of small order in n compared with $\frac{1}{2}nQ$ and may therefore be neglected in deriving estimating equations for the β 's. For completeness, however, the evaluation of $|B|$ was investigated for a second-order process and in the Appendix it is shown to have the limiting value $1/[(1 - \beta_2)^2 \{(1 + \beta_2)^2 - \beta_1^2\}]$ for large k . Neglecting $|B|$, on differentiating Q with respect to $\beta_1, \dots, \beta_h$ and equating the derivatives to zero we obtain the estimators $b_1, \dots, b_h$ of $\beta_1, \dots, \beta_h$ as the solution of the linear equations

$$\begin{bmatrix} \sum_{i=0}^k a_i^2 & \sum_{i=0}^{k-1} a_i a_{i+1} & \sum_{i=0}^{k-2} a_i a_{i+2} & \dots & \sum_{i=0}^{k-h+1} a_i a_{i+h-1} \\ \sum_{i=0}^{k-1} a_i a_{i+1} & \sum_{i=0}^k a_i^2 & & & \\ \vdots & & \ddots & & \\ \sum_{i=0}^{k-h+1} a_i a_{i+h-1} & \dots & \dots & \dots & \sum_{i=0}^k a_i^2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_h \end{bmatrix} = - \begin{bmatrix} \sum_{i=0}^{k-1} a_i a_{i+1} \\ \sum_{i=0}^{k-2} a_i a_{i+2} \\ \vdots \\ \sum_{i=0}^{k-h} a_i a_{i+h} \end{bmatrix}. \quad (15)$$

The asymptotic variance matrix of $b_1, \dots, b_h$ is approximately $\frac{1}{n} E \left[\frac{\partial^2 Q}{\partial \beta_i \partial \beta_j} \right]^{-1}$, which to the first order in n is U/n , where

$$U^{-1} = \begin{bmatrix} \sum_{i=0}^k \alpha_i^2 & \sum_{i=0}^{k-1} \alpha_i \alpha_{i+1} & \dots & \sum_{i=0}^{k-h+1} \alpha_i \alpha_{i+h-1} \\ \vdots & \ddots & & \\ \sum_{i=0}^{k-h+1} \alpha_i \alpha_{i+h-1} & \dots & \dots & \sum_{i=0}^k \alpha_i^2 \end{bmatrix}.$$

Here, $\sum_{i=1}^{k-s} \alpha_i \alpha_{i+s}$ is the coefficient of z^s in the expansion of

$$(1 + \alpha_1 z + \dots + \alpha_k z^k)(1 + \alpha_1 z^{-1} + \dots + \alpha_k z^{-k}),$$

which for large k is nearly equal to the coefficient of z^s in the expansion of

$$(1 + \beta_1 z + \dots + \beta_h z^h)^{-1} (1 + \beta_1 z^{-1} + \dots + \beta_h z^{-h})^{-1}.$$

This is equal to the covariance of y_t and y_{t+s} in the autoregressive series generated by

$$y_t + \beta_1 y_{t-1} + \dots + \beta_h y_{t-h} = \xi_t, \quad (16)$$

where $\{\xi_t\}$ has zero mean and unit variance. Thus U^{-1} is the variance matrix of h successive observations of the series (16). Its inverse, U , may be obtained by means of the following considerations.

Let Y_1 denote the (column) vector $\{y_1, \dots, y_h\}$ and Y_2 the vector $\{y_{h+1}, \dots, y_{2h}\}$. The unconditional density of Y_1 and the conditional density of Y_2 given Y_1 are

$$K_1 \exp\left(-\frac{1}{2} Y_1' U Y_1\right) \quad \text{and} \quad K_2 \exp\left\{-\frac{1}{2} \sum_{t=h+1}^{2h} (y_t + \beta_1 y_{t-1} + \dots + \beta_h y_{t-h})^2\right\},$$

respectively, where K_1 and K_2 are suitable constants. Thus the unconditional density of $y_1, \dots, y_{2h}$ is

$$K_1 K_2 \exp\left[-\frac{1}{2} \left\{ Y_1' U Y_1 + \sum_{t=h+1}^{2h} (y_t + \beta_1 y_{t-1} + \dots + \beta_h y_{t-h})^2 \right\}\right]. \quad (17)$$

Similarly, the density of Y_2 and the conditional density of Y_1 given Y_2 are, since a stationary time series is symmetric with regard to direction along the time axis,

$$K_1 \exp\left(-\frac{1}{2} Y_2' U Y_2\right) \quad \text{and} \quad K_2 \exp\left\{-\frac{1}{2} \sum_{t=1}^h (y_t + \beta_1 y_{t+1} + \dots + \beta_h y_{t+h})^2\right\}.$$

Thus the unconditional density of $y_1, \dots, y_{2h}$ is

$$K_1 K_2 \exp\left[-\frac{1}{2} \left\{ \sum_{t=1}^h (y_t + \beta_1 y_{t+1} + \dots + \beta_h y_{t+h})^2 + Y_2' U Y_2 \right\}\right]. \quad (18)$$

(17) and (18) must be identical since they represent the same density. Equating the first h rows and columns of the matrices of the quadratic forms in (17) and (18) we have

$$\begin{bmatrix} u_{11} + \beta_h^2 & u_{12} + \beta_{h-1}\beta_h & \dots & u_{1h} + \beta_1\beta_h \\ u_{12} + \beta_{h-1}\beta_h & u_{22} + \beta_{h-1}^2 + \beta_h^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ u_{1h} + \beta_1\beta_h & \dots & \dots & \beta_1^2 + \dots + \beta_h^2 \end{bmatrix} = \begin{bmatrix} 1 & \beta_1 & \dots & \beta_{h-1} \\ \beta_1 & 1 + \beta_1^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \beta_{h-1} & \dots & \dots & 1 + \beta_1^2 + \dots + \beta_{h-1}^2 \end{bmatrix}.$$

On subtraction, we find

$$U = \begin{bmatrix} 1 - \beta_h^2 & \beta_1 - \beta_{h-1}\beta_h & & & & \\ \beta_1 - \beta_{h-1}\beta_h & 1 + \beta_1^2 - \beta_{h-1}^2 - \beta_h^2 & & & & \\ \beta_2 - \beta_{h-2}\beta_h & \beta_1 + \beta_1\beta_2 - \beta_{h-2}\beta_{h-1} - \beta_{h-1}\beta_h & & & & \\ \vdots & & \ddots & & & \\ \beta_{h-1} - \beta_1\beta_h & & & & & \end{bmatrix}. \quad (19)$$

Values of U for $h = 1, 2, 3$ are

$$\begin{array}{ccc}
 h = 1 & h = 2 & h = 3 \\
 [1 - \beta_1^2]; & \begin{bmatrix} 1 - \beta_2^2 & \beta_1 - \beta_1\beta_2 \\ \beta_1 - \beta_1\beta_2 & 1 - \beta_2^2 \end{bmatrix}; & \begin{bmatrix} 1 - \beta_3^2 & \beta_1 - \beta_2\beta_3 & \beta_2 - \beta_1\beta_3 \\ \beta_1 - \beta_2\beta_3 & 1 + \beta_1^2 - \beta_2^2 - \beta_3^2 & \beta_1 - \beta_2\beta_3 \\ \beta_2 - \beta_1\beta_3 & \beta_1 - \beta_2\beta_3 & 1 - \beta_3^2 \end{bmatrix}.
 \end{array}$$

6. REMARKS ON THE ASSUMPTIONS

The assumption of normality in the distribution of the ϵ 's in models (1) and (13) was made for convenience and can be dropped without much loss. The justification is that Mann & Wald (1943) in deriving the asymptotically normal distribution of $a_1, \dots, a_k$ assume only that the ϵ 's are independent and identically distributed variables with a distribution having zero mean and finite moments of all orders. On these wider assumptions the estimates b and $b_1, \dots, b_k$ derived above are still consistent in the limit with limiting variance and variance matrix given by (12) and U/n , where U is given by (19), respectively. What we lose by adopting the wider assumptions is the assurance that the estimators are efficient. However, it is possible that the estimators could be proved to be efficient in the class of estimators based on the sample serial correlation coefficients. This possibility ensues from the asymptotic normality of the serial correlation coefficients of a moving-average process taken together with Bartlett's (1946) result that first and second moments of the sample serial correlation coefficients are independent of the parent distribution of the ϵ 's.

The assumption that $|\beta| < 1$ in model (1) is less restrictive than might appear at first sight. It is customary to impose the requirement $|\beta| \leq 1$, since otherwise there is an essential indeterminacy arising from the fact that we cannot distinguish from the study of observed values between observations generated by the model with $\beta = \beta_0$ and observations generated by the model with $\beta = 1/\beta_0$. The indeterminacy disappears if we follow the convention $|\beta| \leq 1$. If $|\beta| = 1$, the method proposed breaks down since the autoregressive representation does not converge. However, other methods based on series expansions also break down, e.g. methods based on the form of the likelihood given in (2). The point is not of much practical importance since such cases are hardly likely to arise. Similar remarks apply to the assumption for model (13) that the roots of $x^h + \beta_1 x^{h-1} + \dots + \beta_h = 0$ have modulus less than one.

It was assumed in §3 that the estimates $a_1, \dots, a_k$ of $\alpha_1, \dots, \alpha_k$ are calculated by least squares. Many statisticians will, however, prefer to work with estimates $a'_1, \dots, a'_k$ calculated from the sample serial correlation coefficients $r_1, \dots, r_k$ using the relations

$$\begin{array}{rcl}
 a'_1 + r_1 a'_2 + \dots + r_{k-1} a'_k + r_1 & = & 0 \\
 r_1 a'_1 + a'_2 + \dots + r_{k-2} a'_k + r_2 & = & 0 \\
 \vdots & & \vdots \\
 r_{k-1} a'_1 + r_{k-2} a'_2 + \dots + a'_k + r_k & = & 0.
 \end{array}$$

Since there is no difference in asymptotic behaviour between $a'_1, \dots, a'_k$ and the least-squares estimates $a_1, \dots, a_k$, the asymptotic theory given above holds for estimates based on $a'_1, \dots, a'_k$ as well as for estimates based on $a_1, \dots, a_k$.

7. TESTS OF SIGNIFICANCE AND TESTS OF FIT

The above results furnish a basis for the construction of large-sample tests of significance. Thus to test the hypothesis $\beta = \beta_0$ in (1) we calculate b from (7) and test

$$z = \sqrt{n} (b - \beta_0) (1 - \beta_0^2)^{-\frac{1}{2}}$$

as a standard normal deviate. Similarly, to test the hypothesis $\beta_i = \beta_{0i}$ ($i = 1, \dots, h$) in (13) we test $w = n \sum_{i=1}^h \sum_{j=1}^h u^{ij} (b_i - \beta_{0i}) (b_j - \beta_{0j})$ as a χ^2 variable with h degrees of freedom, where $[u^{ij}] = U^{-1}$ and U is calculated from $\beta_{01}, \dots, \beta_{0h}$ using (19).

A test of goodness-of-fit of model (1) is obtained by noting from (5) and (6) that

$$nQ = n \left\{ (1 + \beta^2) \sum_{i=0}^k a_i^2 + 2\beta \sum_{i=0}^{k-1} a_i a_{i+1} - 1 \right\}$$

is approximately distributed on χ^2 with k degrees of freedom. It may be verified by substitution from (7) that Q can be partitioned in the form

$$Q = (1 - b^2) \sum_{i=0}^k a_i^2 - 1 + (b - \beta)^2 \sum_{i=0}^k a_i^2.$$

Asymptotically, the term $n(b - \beta)^2 \sum a_i^2$ is equivalent to the regression sum of squares in a linear regression model, while the remainder is equivalent to the residual sum of squares. Thus the goodness-of-fit of the model may be examined by testing

$$n\hat{Q} = n \left\{ (1 - b^2) \sum_{i=0}^k a_i^2 - 1 \right\} \quad (20)$$

as a χ^2 variable with $k - 1$ degrees of freedom.

The test for the general model (13) is obtained by recalling that $b_1, \dots, b_h$ are obtained by minimizing the expression Q defined by (14) with respect to $\beta_1, \dots, \beta_h$. Substituting $b_1, \dots, b_h$ for $\beta_1, \dots, \beta_h$ in (14) and using the relations (15) we find for the minimum value of Q ,

$$\hat{Q} = \sum_{i=0}^k a_i^2 + \sum_{j=1}^h b_j \sum_{i=0}^{k-j} a_i a_{i+j} - 1. \quad (21)$$

The goodness-of-fit of the model is tested by treating $n\hat{Q}$ as a χ^2 variable with $k - h$ degrees of freedom.

This test of fit is clearly a good deal simpler to use than that suggested by Wold (1949).

8. SOME NUMERICAL RESULTS

First-order moving-averages have been fitted to twenty series each containing 100 observations generated by the model

$$x_t = \epsilon_t + \frac{1}{2}\epsilon_{t-1}.$$

The $\{\epsilon_t\}$ were pseudo-random normal deviates calculated on the English Electric DEUCE calculator. Autoregressive models of order five were fitted to each series. Denoting the resulting coefficients by $a_1, \dots, a_5$ the moving-average parameter was then estimated by the formula

$$b = - \frac{a_1 + a_1 a_2 + \dots + a_1 a_5}{1 + a_1^2 + \dots + a_5^2}.$$

For each series the simple estimator c obtained from the first serial correlation r_1 was also calculated. This is given by the formula

$$c/(1+c^2) = r_1,$$

taking the root having modulus less than one. Where the roots were imaginary c was taken to be one. The results are given in Table 1.

Table 1

Series	b	c	Series	b	c
1	0.4302	0.4085	11	0.5199	0.5916
2	.3639	.4966	12	.5304	.4418
3	.3851	.3509	13	.5278	.4467
4	.4751	.4579	14	.4856	.5424
5	.4298	.3863	15	.3141	.3524
6	0.5466	0.6665	16	0.3380	0.2670
7	.5326	.6106	17	.4005	.4648
8	.4156	.6212	18	.5528	1.0000
9	.4926	1.0000	19	.3993	0.3649
10	.4142	0.2839	20	.5069	.4984

The sample means and variances are:

	b	c
Mean	0.4531	0.5126
Variance	0.00536	0.0398
S.E. of mean	0.0164	0.0446

The claim for the efficiency of b is well supported by these results. The observed variance, namely 0.00536, is actually less than the theoretical value of $\frac{1}{100}\{1 - (1/2^2)\} = 0.0075$, and is substantially less than the estimated variance of c . A further point in favour of b is that it is closer to the true value of $\frac{1}{2}$ for 16 out of the 20 samples.

On the other hand, it is disappointing that the results for b show such a strong downward bias. The discrepancy between the observed mean of 0.4531 and the true value of $\frac{1}{2}$ is well over twice the estimated standard error. The following remedial measures were tried, but in no case was the improvement satisfactory enough to warrant inclusion of the results: the order of the fitted autoregressive scheme was raised from five to ten; a term was dropped from the denominator of b in the manner referred to at the end of § 3; the term $-\frac{1}{2} \log(1 - \beta^2)$ in the likelihood function was allowed for as described at the end of § 3; Quenouille's method of bias reduction was used (see Durbin, 1959, for details of the method). This matter of bias clearly calls for further consideration.

I am indebted to Professor M. G. Kendall for some valuable corrections to the first draft of the paper and to Miss J. Grahame and Miss J. May for assistance with the computations.

REFERENCES

- BARTLETT, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time series. *J. R. Statist. Soc. (Suppl.)* 8, 27-41.
- CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.
- DIXON, W. J. (1944). Further contributions to the problem of serial correlation. *Ann. Math. Statist.* 15, 119-44.
- DURBIN, J. (1959). A note on Quenouille's method of bias reduction. *Biometrika*, 46, 477-80.
- MANN, H. B. & WALD, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica*, 11, 173-220.
- WHITTLE, P. (1951). *Hypothesis Testing in Time Series*. Uppsala: Almqvist and Wiksells.
- WHITTLE, P. (1953). Estimation and information in stationary time series. *Ark. Mat.* 2, 423-34.
- WOLD, H. (1938). *A Study in the Analysis of Stationary Time Series*, 1st ed. Uppsala: Almqvist and Wiksells.
- WOLD, H. (1949). A large-sample test for moving-averages. *J. R. Statist. Soc. B*, 11, 297-305.

APPENDIX

Evaluation of the limiting covariance determinant of the second-order moving-average process

Let D_n be the n th-order determinant

$$\begin{vmatrix} a & b & c & 0 & 0 & \dots & \dots & \dots & 0 \\ b & a & b & c & 0 & & & & \vdots \\ c & b & a & b & c & & & & \vdots \\ 0 & c & b & a & b & & & & \vdots \\ \vdots & & & & \ddots & & & & \vdots \\ \vdots & & & & & \ddots & & & \vdots \\ \vdots & & & & & & \ddots & & \vdots \\ \vdots & & & & & & & a & b \\ \vdots & & & & & & & b & a \\ 0 & \dots & \dots & \dots & \dots & 0 & c & b & a \end{vmatrix}.$$

We proceed to calculate $\lim_{n \rightarrow \infty} D_n$ for the case where $a = 1 + \beta_1^2 + \beta_2^2$, $b = \beta_1 + \beta_1\beta_2$, $c = \beta_2$.

Expanding by the first row we obtain after some reduction the fifth-order difference equation

$$\{-1 + (a-c)z + (ac-b^2)z^2 + c(b^2-ac)z^3 + c^3(c-a)z^4 + c^5z^5\} D_n = 0, \quad (A1)$$

where $z^r D_n$ denotes D_{n-r} . On substitution for a, b, c we find that (A1) factorizes in the form

$$(z-1)(\beta_2 z-1)(\beta_2^2 z-1)\{\beta_2^2 z^2 + (2\beta_2 - \beta_1^2)z + 1\} D_n = 0. \quad (A2)$$

Let

$$\{\beta_2^2 z^2 + (2\beta_2 - \beta_1^2)z + 1\} D_n = C_n. \quad (A3)$$

Then C_n satisfies the difference equation

$$(z-1)(\beta_2 z-1)(\beta_2^2 z-1)C_n = 0.$$

On examination it turns out that equation (A1) is satisfied for $n \geq 2$ provided we take $D_0 = 1$, $D_{-1} = 0$, $D_{-2} = 0$. Substituting in (A2) we have

$$\left. \begin{aligned} C_0 &= 1, \\ C_1 &= (1 + \beta_2)^2, \\ C_2 &= (1 + \beta_2^2)(1 + \beta_2)^2 + \beta_2^2, \\ C_n &= \lambda_1 + \lambda_2 \beta_2^n + \lambda_3 \beta_2^{2n}. \end{aligned} \right\} \quad (A4)$$

The general solution of (A3) is

When specifying the moving-average model we stipulated that $x^2 + \beta_1 x + \beta_2 = 0$ has roots with modulus less than one. It follows that $|\beta_2| < 1$. Consequently $\lim_{n \rightarrow \infty} C_n = \lambda_1$. Substituting in (A4) we find $\lambda_1 = (1 - \beta_2)^{-2}$.

From (A2) we have that

$$\{\beta_2^2 + (2\beta_2 - \beta_1^2)z + 1\} \lim_{n \rightarrow \infty} D_n = \lim_{n \rightarrow \infty} C_n.$$

Consequently,
$$\lim_{n \rightarrow \infty} D_n = \frac{1}{(1 - \beta_2)^2 \{(1 + \beta_2)^2 - \beta_1^2\}}. \quad (\text{A } 5)$$

It is remarkable that this limiting value is the exact reciprocal of the very similar determinant

$$H_n = \begin{vmatrix} 1 & \beta_1 & \beta_2 & 0 & 0 & \dots & 0 \\ \beta_1 & 1 + \beta_1^2 & \beta_1(1 + \beta_2) & \beta_2 & 0 & & \vdots \\ \beta_2 & \beta_1(1 + \beta_2) & 1 + \beta_1^2 + \beta_2^2 & \beta_1(1 + \beta_2) & & & \vdots \\ 0 & \beta_2 & \beta_1(1 + \beta_2) & 1 + \beta_1^2 + \beta_2^2 & & & \vdots \\ \vdots & & & & \ddots & & \vdots \\ 0 & \dots & \dots & \dots & \dots & 1 + \beta_1^2 & \beta_1 \\ & & & & & \beta_1 & 1 \end{vmatrix} \quad (n \geq 5),$$

which differs from D_n only in the first and last two rows and columns. H_n is the determinant of the reciprocal of the variance matrix of n successive observations of the autoregressive series

$$x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} = \epsilon_t,$$

where ϵ_t has unit variance. Thus the limiting covariance determinant of n successive values of a second-order moving-average series is the same as the covariance determinant of n observations of a second-order autoregressive series with the same coefficients and the same residual variance. It is likely that this result is generally true and it would be interesting to see a general proof of it.

THE COMPLETE AMALGAMATION INTO BLOCKS, BY WEIGHTED MEANS, OF A FINITE SET OF REAL NUMBERS

BY R. E. MILES

Statistical Laboratory, University of Cambridge

1. INTRODUCTION AND SUMMARY

The problem discussed below arose in Bartholomew (1959*a*), in which is proposed a test of homogeneity of 'ordered' alternatives.

On p. 40 of this paper Bartholomew defined the probability $P(l, k; a_1, a_2, \dots, a_k)$ and proceeded to determine its values for $1 \leq l \leq k \leq 4$ and general a_i ($i = 1, \dots, k$).

In the special, but important, case of equal weights ($a_1 = a_2 = \dots = a_k$) it may readily be verified that $P(l, k; a_1, a_2, \dots, a_k)$ is independent of the common value of the a_i , and thus may be written $\hat{P}(l, k)$, say; this notation departs slightly from Bartholomew's in the use of a circumflex, but this is done to avoid confusion in the Appendix. Bartholomew determined values of $\hat{P}(l, k)$ for $1 \leq l \leq k \leq 5$ and, on the basis of these results, conjectured on p. 43 the recurrence relations

$$\hat{P}(l, k) = \frac{1}{l} \sum_{j=l-1}^{k-1} \hat{P}(1, k-j) \hat{P}(l-1, j) \quad (2 \leq l \leq k), \quad (*)$$

(**)

where
$$\hat{P}(1, k) = \frac{1}{k}.$$

We now make a brief survey of the present paper.

Summary

No knowledge is pre-supposed of Bartholomew (1959*a*), and the notation and terminology differ to some extent. In § 2 an ' n -collection' $[n]$ is defined as a set of n real numbers, called 'ordinates', and n positive real numbers, called 'weights',[†] satisfying a certain condition $\mathcal{C}$. It is then possible to define a 'random array' of $[n]$. The 'complete amalgamation', by the amalgamation process $\mathcal{A}_1$, of an array into 'blocks' by decreasing weighted means—a generalization of Bartholomew's amalgamation process—is described in § 3; with the aid of an equivalent amalgamation process $\mathcal{A}_2$, the complete amalgamation into blocks by $\mathcal{A}_1$ is proved to be unique (Theorem 1). In § 4, $P_{[n]}(k, n)$ is defined as the probability that a random array of $[n]$ yields k blocks on complete amalgamation. There follows the interesting main result of the paper (Theorem 2), viz. $P_{[n]}(k, n) = P(k, n)$, independent of $[n]$, and recurrence relations (10)–(13) are given which determine the $P(k, n)$; in the Corollary to Theorem 2 the ordinates are taken as a sample from a continuous probability distribution, and this provides the final link with Bartholomew (1959*a*). The generating functions of $P(k, n)$ are determined in § 5 and, making use of one of these expressions, we obtain in § 6 the formula

$$P(k, n) = \frac{1}{n!} |S_n^k|,$$

[†] It should be noted that the 'weights' we define have a different meaning from those used by Bartholomew.

where S_n^k are the Stirling Numbers of the First Kind. General and limiting properties, and a table of values, of $P(k, n)$ are given in § 7. Finally, in § 8 two related problems are discussed which both yield an identical probability distribution.

The conjecture (*)-(**) is in fact true. This is proved in the Appendix where we obtain the result

$$\hat{P}(k, n) = P(k, n) = \frac{1}{n!} |S_n^k|.$$

On account of this, Bartholomew's test in the case of equal weights a_i may be completely specified. Further details arising from this development are included in a second paper—Bartholomew (1959b).

2. DEFINITIONS AND NOTATION

We consider a set of n real numbers $X_1, \dots, X_n$ —the 'ordinates', and a set of n positive real numbers $W_1, \dots, W_n$ —the 'weights'.

We define a condition $\mathcal{C}$ on the weighted means of the ordinates with respect to the weights as follows:

$$\mathcal{C}: \frac{W_{\rho'_1} X_{\rho_1} + \dots + W_{\rho'_r} X_{\rho_r}}{W_{\rho'_1} + \dots + W_{\rho'_r}} \neq \frac{W_{\rho'_{r+1}} X_{\rho_{r+1}} + \dots + W_{\rho'_{r+s}} X_{\rho_{r+s}}}{W_{\rho'_{r+1}} + \dots + W_{\rho'_{r+s}}}$$

for any two $(r+s)$ -permutations $\rho(\rho_1 \dots \rho_{r+s})$, $\rho'(\rho'_1 \dots \rho'_{r+s})$ of the integers $1, \dots, n$, where $1 \leq r, s \leq n-1$ and $2 \leq r+s \leq n$.

Imposing $\mathcal{C}$ in what follows enables us to use strict inequalities ($>$, $<$) instead of weak ones ($\geq$, $\leq$). In particular, taking $r = s = 1$, we see that $\mathcal{C}$ requires that $X_i \neq X_j$ if $i \neq j$, i.e. the X_i are n different real numbers; however, $\mathcal{C}$ places no corresponding restriction on the W_i .

We define an ' n -collection' as a set of n ordinates and n weights satisfying $\mathcal{C}$, and denote it by $[n]$. By an ' n -array' is meant a $2 \times n$ matrix whose elements constitute an n -collection—the first row being the ordinates, and the second row the weights, in some order. $[n]$ thus gives rise to $(n!)^2$ different n -arrays—the 'arrays of $[n]$ '. By an r -subarray of an n -array ($1 \leq r \leq n$) is meant an r -array obtained by deleting columns of the n -array so that only r adjacent columns remain. $\mathcal{C}$ holds *a fortiori* for the elements of subarrays. (Dropping the prefix, the terms collection, array and subarray have obvious meanings.)

We select independently, at random, two permutations $\sigma(\sigma_1 \dots \sigma_n)$, $\sigma'(\sigma'_1 \dots \sigma'_n)$ of the integers $1, \dots, n$. There are $(n!)^2$ ways in which we can do this, all equally probable, and so each ordered permutation-pair (σ, σ') has probability $1/(n!)^2$ of being selected.

We call

$$S_{n,n}^{\sigma, \sigma'} = \begin{pmatrix} X_{\sigma_1} \dots X_{\sigma_n} \\ W_{\sigma'_1} \dots W_{\sigma'_n} \end{pmatrix},$$

the n -array corresponding to our random choice, a 'random array' (R.A.) of $[n]$.

We write

$$\begin{aligned} Z_j^i &= W_{\sigma_i} X_{\sigma_i} + W_{\sigma'_{i+1}} X_{\sigma'_{i+1}} + \dots + W_{\sigma'_j} X_{\sigma'_j}, \\ W_j^i &= W_{\sigma_i} + W_{\sigma'_{i+1}} + \dots + W_{\sigma'_j}, \\ X_j^i &= \frac{Z_j^i}{W_j^i} \quad (\text{a 'weighted mean'}), \end{aligned}$$

and Y_j^i as the 2×1 matrix $\begin{pmatrix} X_j^i \\ W_j^i \end{pmatrix}$, where $1 \leq i \leq j \leq n$. It will be noticed that Y_j^i is a 1-array, and that we may write

$$S_{n,n}^{\sigma, \sigma'} = (Y_1^1 \dots Y_n^n). \quad (1)$$

3. THE COMPLETE AMALGAMATION OF AN ARRAY

$S_{n,n}^{\sigma,\sigma'}$ is taken as the array.

The amalgamation process $\mathcal{A}_1$. We operate stage by stage on $S_{n,n}^{\sigma,\sigma'}$ as follows. In each stage, having started afresh with an r -array ($2 \leq r \leq n$), either the process terminates or we finish with a new $(r-1)$ -array. Thus, after the $(n-r)$ th stage, we have the r -array

$$S_{n,r}^{\sigma,\sigma'} = (Y_{r_1}^{r_0+1} \dots Y_{r_r}^{r_{r-1}+1}), \quad \text{where } r_0 \equiv 0, r_r \equiv n. \quad (2)$$

If $X_{r_1}^{r_0+1} > \dots > X_{r_r}^{r_{r-1}+1}$ we proceed no further, the process being said to have terminated. Otherwise there are one or more 2-subarrays $(Y_{r_i}^{r_{i-1}+1} Y_{r_{i+1}}^{r_i+1})$, where $1 \leq i \leq r-1$, in $S_{n,r}^{\sigma,\sigma'}$ such that $X_{r_i}^{r_{i-1}+1} < X_{r_{i+1}}^{r_i+1}$. In this case we choose quite freely any one of these subarrays and replace it by the 1-subarray $Y_{r_{i+1}}^{r_i+1}$, thereby obtaining the $(r-1)$ -array $S_{n,r-1}^{\sigma,\sigma'}$. We thus obtain successively the arrays $S_{n,n}^{\sigma,\sigma'}, S_{n,n-1}^{\sigma,\sigma'}, \dots, S_{n,k}^{\sigma,\sigma'}$ the process terminating after $n-k$ stages, where $1 \leq k \leq n$; the final k -array $S_{n,k}^{\sigma,\sigma'}$ is such that $X_{k_1}^1 > \dots > X_n^{k_{k-1}+1}$. It should be noted that the process automatically terminates if $S_{n,1}^{\sigma,\sigma'}$ is reached.

We observe that $S_{n,r-1}^{\sigma,\sigma'}$ has the same basic form as $S_{n,r}^{\sigma,\sigma'}$, as written on the right-hand side of (2); by (1), $S_{n,n}^{\sigma,\sigma'}$ also has this property. Hence, by induction, so do all the $S_{n,s}^{\sigma,\sigma'}$ ($k \leq s \leq n$). This ensures that the process is a valid one and may always be carried out as described.

We say that the R.A. $S_{n,n}^{\sigma,\sigma'}$ of $[n]$ has yielded k 'blocks' $Y_{k_1}^1, \dots, Y_n^{k_{k-1}+1}$ in this 'complete amalgamation' (complete amal.) by $\mathcal{A}_1$.

It is seen immediately, from the nature of $\mathcal{A}_1$, that $S_{n,n-1}^{\sigma,\sigma'}, \dots, S_{n,k+1}^{\sigma,\sigma'}$ are not, in general, unique.

However, we now prove that k and $S_{n,k}^{\sigma,\sigma'}$ are unique. This we do by first defining an alternative amalgamation process $\mathcal{A}_2$ which gives a unique complete amal. Then we demonstrate (Theorem 1) the exact equivalence of a complete amal. by $\mathcal{A}_1$ with the complete amal. by $\mathcal{A}_2$.

The amalgamation process $\mathcal{A}_2$. We take as first block $Y_{l_1}^1$, where $X_{l_1}^1 = \max_{l=1, \dots, n} X_l^1$; $Y_{l_1}^{l_1+1}$ as second block, where $X_{l_1}^{l_1+1} = \max_{l=(l_1+1), \dots, n} X_l^{l_1+1}$; and so on. The process terminates after l steps, where $1 \leq l \leq n$, giving the complete amal. of $S_{n,n}^{\sigma,\sigma'}$ into l blocks by $\mathcal{A}_2$.

THEOREM 1. Any complete amal. of an array by $\mathcal{A}_1$ is identical with its unique complete amal. by $\mathcal{A}_2$.

Proof. We first show, as a consequence of $X_{l_1}^1$ being $\max_{l=1, \dots, n} X_l^1$, that

$$X_{l_1}^i > X_{l_1}^1 > X_j^{l_1+1} \quad (3)$$

for all i, j such that $2 \leq i \leq l_1$ and $l_1+1 \leq j \leq n$. The first inequality is excluded if $l_1 = 1$.

For suppose the first part of (3) did not hold. Then

$$Z_{l_1}^i < W_{l_1}^i X_{l_1}^1. \quad (4)$$

$$Z_{l_1-1}^1 < W_{l_1-1}^1 X_{l_1}^1. \quad (5)$$

By definition

Adding (4) and (5), we obtain $Z_{l_1}^1 < W_{l_1}^1 X_{l_1}^1$ or $X_{l_1}^1 < X_{l_1}^1$ —a contradiction.

Now suppose the second part of (3) did not hold. Then

$$W_j^{l_1+1} X_{l_1}^1 < Z_j^{l_1+1}. \quad (6)$$

$$W_{l_1}^1 X_{l_1}^1 = Z_{l_1}^1. \quad (7)$$

By definition

Adding (6) and (7), we obtain $W_j^1 X_{l_1}^1 < Z_j^1$ or $X_{l_1}^1 < X_j^1$ —another contradiction.

The proof of (3) is thus completed.

It immediately follows that no amalgamation is possible in $\mathcal{A}_1$ between any ordinate of the type $X_{l_i}^i$ with one of the type $X_{l_i}^{i+1}$. Thus $\mathcal{A}_1$ is at least as 'fine' as $\mathcal{A}_2$. Hence it now remains only to prove conversely that $\mathcal{A}_2$ is at least as fine as $\mathcal{A}_1$, i.e. that in a complete amal. by $\mathcal{A}_1$ the l_1 -array $(Y_1^1 \dots Y_{l_1}^1)$ cannot be separated into more than one block—into t blocks, say, where $t \geq 2$.

For, suppose this were possible, these blocks being $Y_{l_i}^1, \dots, Y_{l_i}^{t-1+1}$. Then

$$X_{l_i}^{t-1+1} > X_{l_i}^1 \quad (8)$$

by the first inequality of (3). Also, since this is a complete amal. by $\mathcal{A}_1$,

$$X_{l_i}^1 > \dots > X_{l_i}^{t-1+1}. \quad (9)$$

(8) and (9) together imply that $X_{l_i}^1 > X_{l_i}^1$, a contradiction.

Hence the first block in all complete amals. by $\mathcal{A}_1$ is the same as the first block in the complete amal. by $\mathcal{A}_2$. We may now similarly prove that the second block has the same property. And so on.

We may define an amalgamation process $\mathcal{A}_3$ which is the 'reverse' of $\mathcal{A}_2$.

The amalgamation process $\mathcal{A}_3$. We take as last block $Y_n^{m_{m-1}+1}$, where $X_n^{m_{m-1}+1} = \min_{m=1, \dots, n} X_n^m$; $Y_{m_{m-1}}^{m_{m-1}+1}$ as penultimate block, where $X_{m_{m-1}}^{m_{m-1}+1} = \min_{m=1, \dots, m_{m-1}} X_{m_{m-1}}^m$; and so on. This process terminates after m steps, where $1 \leq m \leq n$, giving the complete amal. of $S_{n,n}^{\sigma, \sigma'}$ into m blocks by $\mathcal{A}_3$.

We may prove the theorem resulting from the replacement of $\mathcal{A}_2$ by $\mathcal{A}_3$ in the statement of Theorem 1 by a similar method.

Thus we may now speak without ambiguity of the 'complete amal. of an array into blocks', without specifying which of the processes $\mathcal{A}_1$, $\mathcal{A}_2$ or $\mathcal{A}_3$ has been employed.

4. THE PROBABILITIES $P(k, n)$

Let $*$ be some condition we may impose on a R.A. of $[n]$. We define

$$\begin{aligned} P_{[n]}(k, n \mid *) &= \text{prob. (R.A. of } [n] \text{ yields } k \text{ blocks} \mid \text{R.A. satisfies } *) \\ &= \frac{\text{no. of R.A.'s of } [n] \text{ satisfying } * \text{ and yielding } k \text{ blocks}}{\text{no. of R.A.'s of } [n] \text{ satisfying } *} \end{aligned}$$

and $P_{[n]}(k, n)$ as the corresponding prob. with no condition on the R.A.'s.

We notice that $P_{[n]}(k, n \mid *)$, $P_{[n]}(k, n)$ are only defined for integers k, n such that $1 \leq k \leq n$.

THEOREM 2. $P_{[n]}(k, n)$ is independent of the n -collection $[n]$, and thus may be written $P(k, n)$. The $P(k, n)$ are determined by

$$P(1, 1) = 1 \quad (10)$$

and the three recurrence relations (for $n \geq 2$)

$$P(1, n) = \frac{n-1}{n} P(1, n-1), \quad (11)$$

$$P(k, n) = \frac{1}{n} [P(k-1, n-1) + (n-1) P(k, n-1)] \quad (2 \leq k \leq n-1), \quad (12)$$

$$\text{and} \quad P(n, n) = \frac{1}{n} P(n-1, n-1). \quad (13)$$

Proof. (10) is immediate.

To complete the proof of Theorem 2, we proceed by induction. Let $\mathcal{P}_r$ denote the proposition $\{P_{[r]}(k, r) \text{ is independent of } [r] \text{ for } 1 \leq k \leq r, \text{ and thus equals } P(k, r)\}$. Let us suppose that $\mathcal{P}_{n-1}$ is true.

Let $X_{(i)} = \max_{i=1, \dots, n} X_i$. $X_{(i)}$ occurs, with equal probability $1/n$, in each of the 'positions' (synonymous with columns of a matrix) $1, \dots, n$ of a R.A. of $[n]$.

We subdivide the arrays of $[n]$ into 'classes' $\alpha_{i'}, \beta_{j'i'}$, such that each of the $(n!)^2$ arrays of $[n]$ belongs to one, and only one, class. We say a R.A. belongs to $\alpha_{i'} \{i' = 1, \dots, n\}$ if the 1-subarray $\left(\begin{smallmatrix} X_{(i)} \\ W_{i'} \end{smallmatrix}\right)$ occurs in position 1 of the array, and to $\beta_{j'i'} \{j = 1, \dots, (i) - 1, (i) + 1, \dots, n; i', j' = 1, \dots, n, \text{ where } i' \neq j'\}$ if the 2-subarray $\left(\begin{smallmatrix} X_j & X_{(i)} \\ W_{j'} & W_{i'} \end{smallmatrix}\right)$ occurs in any of the positions $(1, 2), (2, 3), \dots, (n-1, n)$ of the array. (We emphasize that a subarray of an array contains *adjacent* columns of the array.)

It is easily seen that the above classes are mutually exclusive, and together contain all the $(n!)^2$ arrays of $[n]$.

The classes α, β have this same property if we write

$$\alpha = \sum_{i'=1}^n \alpha_{i'} \quad \text{and} \quad \beta = \sum_{\text{all admissible } j, j', i'} \beta_{j'i'}$$

We now consider these classes individually.

$\alpha_{i'}$. In a complete amal. the 1-subarray $\left(\begin{smallmatrix} X_{(i)} \\ W_{i'} \end{smallmatrix}\right)$ will obviously be the first block. If in each member of $\alpha_{i'}$ we delete this subarray from position 1, we obtain *every one* of the arrays of the $(n-1)$ -collection $[n-1]_{i'}$ obtained from $[n]$ by omitting the ordinate $X_{(i)}$ and the weight $W_{i'}$. Hence, if an array of $[n-1]_{i'}$ yields $k-1$ blocks, then the corresponding member of $\alpha_{i'}$ yields k blocks; each member of $\alpha_{i'}$ thus yields at least two blocks. But our inductive hypothesis applies to $[n-1]_{i'}$.

(14)

Hence, for $n \geq 2$,

$$P_{[n]}(1, n \mid \text{R.A.} \in \alpha_{i'}) = 0 \quad (15)$$

and

$$P_{[n]}(k, n \mid \text{R.A.} \in \alpha_{i'}) = P_{[n-1]_{i'}}(k-1, n-1) = P(k-1, n-1) \quad (2 \leq k \leq n).$$

$$\text{But} \quad P_{[n]}(k, n \mid \text{R.A.} \in \alpha) = \sum_{i'=1}^n P_{[n]}(k, n \mid \text{R.A.} \in \alpha_{i'}) \text{prob.}(\text{R.A.} \in \alpha_{i'} \mid \text{R.A.} \in \alpha).$$

(16)

Hence, using (14) and (15), we have for $n \geq 2$

$$P_{[n]}(1, n \mid \text{R.A.} \in \alpha) = 0$$

and

$$P_{[n]}(k, n \mid \text{R.A.} \in \alpha) = P(k-1, n-1) \sum_{i'=1}^n \text{prob.}(\text{R.A.} \in \alpha_{i'} \mid \text{R.A.} \in \alpha) = P(k-1, n-1) \quad (2 \leq k \leq n). \quad (17)$$

$\beta_{j'i'}$. Since $X_j < X_{(i)}$, we may by $\mathcal{A}_1$ amalgamate the 2-subarray $\left(\begin{smallmatrix} X_j & X_{(i)} \\ W_{j'} & W_{i'} \end{smallmatrix}\right)$ to form the 1-subarray $\left(\begin{smallmatrix} W_{j'} X_j + W_{i'} X_{(i)} \\ W_{j'} + W_{i'} \end{smallmatrix}\right)$. Thus we cannot obtain n blocks in a complete amal. Moreover, operating in this way on all the members of $\beta_{j'i'}$, we obtain *every one* of the arrays of the

$(n-1)$ -collection $[n-1]_{j',i'}^j$ obtained from $[n]$ by substituting $(W_j, X_j + W_{i'} X_{(i)})$ for X_j and $X_{(i)}$, and $(W_{j'} + W_{i'})$ for $W_{j'}$ and $W_{i'}$, respectively. Hence, if an array of $[n-1]_{j',i'}^j$ yields k blocks, then the corresponding member of $\beta_{j',i'}^j$ also yields k blocks. But our inductive hypothesis applies to $[n-1]_{j',i'}^j$.

Hence, for $n \geq 2$,

$$\begin{aligned} P_{[n]}(k, n \mid \text{R.A.} \in \beta_{j',i'}^j) &= P_{[n-1]_{j',i'}^j}(k, n-1) \\ &= P(k, n-1) \quad (1 \leq k \leq n-1) \end{aligned} \quad (18)$$

$$\text{and} \quad P_{[n]}(n, n \mid \text{R.A.} \in \beta_{j',i'}^j) = 0. \quad (19)$$

But

$$P_{[n]}(k, n \mid \text{R.A.} \in \beta) = \sum_{\text{all admissible}} \sum_{j,j'} \sum_{i'} P_{[n]}(k, n \mid \text{R.A.} \in \beta_{j',i'}^j) \text{prob.}(\text{R.A.} \in \beta_{j',i'}^j \mid \text{R.A.} \in \beta).$$

Hence, using (18) and (19), we have for $n \geq 2$

$$\begin{aligned} P_{[n]}(k, n \mid \text{R.A.} \in \beta) &= P(k, n-1) \sum_{\text{all admissible}} \sum_{j,j'} \sum_{i'} \text{prob.}(\text{R.A.} \in \beta_{j',i'}^j \mid \text{R.A.} \in \beta) \\ &= P(k, n-1) \quad (1 \leq k \leq n-1), \end{aligned} \quad (20)$$

$$\text{and} \quad P_{[n]}(n, n \mid \text{R.A.} \in \beta) = 0. \quad (21)$$

Now

$$P_{[n]}(k, n) = P_{[n]}(k, n \mid \text{R.A.} \in \alpha) \text{prob.}(\text{R.A.} \in \alpha) + P_{[n]}(k, n \mid \text{R.A.} \in \beta) \text{prob.}(\text{R.A.} \in \beta).$$

$$\text{But} \quad \text{prob.}(\text{R.A.} \in \alpha) = \frac{1}{n} \quad \text{and} \quad \text{prob.}(\text{R.A.} \in \beta) = \frac{n-1}{n}.$$

Hence, using (16), (17), (20) and (21), we have

$$P_{[n]}(1, n) = 0 \cdot \frac{1}{n} + P(1, n-1) \frac{n-1}{n}, \quad (22)$$

$$P_{[n]}(k, n) = P(k-1, n-1) \frac{1}{n} + P(k, n-1) \frac{n-1}{n} \quad (2 \leq k \leq n-1), \quad (23)$$

$$\text{and} \quad P_{[n]}(n, n) = P(n-1, n-1) \frac{1}{n} + 0 \cdot \frac{n-1}{n}. \quad (24)$$

Thus we see from (22)–(24) that the truth of $\mathcal{P}_{n-1}$ implies the truth of $\mathcal{P}_n$. But $\mathcal{P}_1$ is true. Hence, by induction, $\mathcal{P}_r$ is true for all positive integers r , and so we may delete the suffix $[n]$ in (22)–(24), thus giving us the recurrence relations (11)–(13).

COROLLARY. All the above theory, including Theorems 1 and 2, applies (with obvious appropriate modification) to the amalgamation of the array $S_n \equiv \begin{pmatrix} x_1 \dots x_n \\ 1 \dots 1 \end{pmatrix}$, where x_i is the i th member of an ordered, i.e. successive, sample of size n from any continuous probability distribution.

Proof. This is because

(a) for ordinates X_i chosen as a finite sample from a continuous probability distribution, and weights W_i chosen as any set of positive real numbers (and hence, in particular, as $W_1 = \dots = W_n = 1$), the condition $\mathcal{C}$ holds with probability 1; and (b) if we are given the members of the sample, but not their order of occurrence, then any one of their $n!$ possible orderings is equally probable to occur.

This work was actuated by an attempt to prove the conjectured recurrence relations (11)–(13) of Theorem 2 for the array S_n of the Corollary. It should be observed that the induction argument cannot be carried through using unit weights, the use of 'generalized' weights being essential. In fact, to prove the Corollary, it would have sufficed in the foregoing theory to take the weights W_i as any set of positive integers which, together with the same ordinates, satisfied $\mathcal{C}$.

5. THE GENERATING FUNCTIONS ASSOCIATED WITH THE $P(k, n)$

As usual, the operations below are merely formal, i.e. they give results which may afterwards be verified; see, for example, Riordan (1958, pp. 19–20).

We define the generating functions $F_k(z)$, $G_n(t)$, and $H(z, t)$ as follows:

$$F_k(z) = \sum_{n=k}^{\infty} P(k, n) z^n,$$

$$G_n(t) = \sum_{k=1}^n P(k, n) t^k$$

$$\text{and } H(z, t) = \sum_{n=1}^{\infty} \sum_{k=1}^n P(k, n) t^k z^n \quad (25)$$

$$= \sum_{k=1}^{\infty} F_k(z) t^k, \quad (26)$$

$$= \sum_{n=1}^{\infty} G_n(t) z^n.$$

Multiplying (11) by tz^n , (12) by $t^k z^n$, (13) by $t^n z^n$, and summing we obtain

$$\frac{\partial H}{\partial z} = \frac{t}{1-z} H,$$

the solution of which is $H = C(1-z)^{-t}$, where C is a constant.

Expanding the right-hand side as a binomial series, (27)

$$H = C \left[1 + tz + \dots + \frac{t(t+1) \dots (t+n-1)}{n!} z^n + \dots \right]. \quad (28)$$

$$G_n(1) = 1. \quad (29)$$

But $\sum_{k=1}^n P(k, n) = 1$, or

$$H(z, t) = (1-z)^{-t}, \quad (30)$$

Thus, by (26)–(28), $C = 1$,

$$G_n(t) = \frac{t(t+1) \dots (t+n-1)}{n!}.$$

and

$$H = \exp \left[t \log \left(\frac{1}{1-z} \right) \right].$$

By (29)

Expanding the right-hand side, this time as an exponential series, (31)

$$H = 1 + t \log \{1/(1-z)\} + \dots + t^k \frac{[\log \{1/(1-z)\}]^k}{k!} + \dots \quad (32)$$

Comparing (25) and (31), it is seen that

$$F_k(z) = \frac{[\log \{1/(1-z)\}]^k}{k!}.$$

(30) and (32) enable us to determine $P(k, n)$ for fixed n and k , respectively.

6. THE RELATIONSHIP OF THE $P(k, n)$ TO THE STIRLING NUMBERS OF THE FIRST KIND, S_n^k

The S_n^k are defined by
$$\sum_{k=1}^n S_n^k t^k = t(t-1) \dots (t-n+1),$$

from which it follows, by substituting $-t$ for t , that

$$\sum_{k=1}^n |S_n^k| t^k = t(t+1) \dots (t+n-1).$$

But, by (30)
$$\sum_{k=1}^n P(k, n) t^k = \frac{t(t+1) \dots (t+n-1)}{n!}.$$

Hence
$$P(k, n) = \frac{1}{n!} |S_n^k|.$$

Since Stirling Numbers are treated rather exhaustively in Jordan (1947, Ch. IV) we shall content ourselves with noting a few of their more important properties in connexion with the $P(k, n)$, giving references where necessary.

7. PROPERTIES OF THE $P(k, n)$

The general solution of equations (10)–(13) is unknown (Jordan, p. 143), but in some cases they can be solved easily;

$$P(1, n) = \frac{1}{n},$$

$$P(n, n) = \frac{1}{n!},$$

(as is otherwise obvious, since there is only one permutation σ such that $X_{\sigma_1} > \dots > X_{\sigma_n}$).

$$P(n-1, n) = \frac{1}{2(n-2)!},$$

$$P(n-2, n) = \frac{1}{8(n-4)!} + \frac{1}{3(n-3)!},$$

and so on (Jordan, p. 149).

We give a table of $n! P(k, n)$ for $1 \leq k \leq n \leq 12$ (Jordan, p. 144).

The mean μ and the variance σ^2 of the number of blocks yielded

$$\mu = \sum_{k=1}^n k P(k, n) = G'_n(1).$$

But

$$\log G_n(t) = \log t + \dots + \log(t+n-1) - \log(n!).$$

Thus

$$\frac{G'_n(t)}{G_n(t)} = \frac{1}{t} + \dots + \frac{1}{t+n-1}. \quad (33)$$

Using (28) and (33)

$$\mu = 1 + \frac{1}{2} + \dots + \frac{1}{n}. \quad (34)$$

The 'mean length of each block', $\bar{l} = \frac{\sum_{k=1}^n (n/k) k P(k, n)}{\sum_{k=1}^n k P(k, n)} = n/\mu$. From (33)

$$\frac{G_n(t) G''_n(t) - [G'_n(t)]^2}{[G_n(t)]^2} = -\frac{1}{t^2} - \dots - \frac{1}{(t+n-1)^2}.$$

Hence

$$\begin{aligned}\sigma^2 &= G_n''(1) + G_n'(1) - [G_n'(1)]^2 \\ &= \left(1 + \frac{1}{2} + \dots + \frac{1}{n}\right) - \left(1 + \frac{1}{2^2} + \dots + \frac{1}{n^2}\right).\end{aligned}\quad (35)$$

Table of $n!P(k, n)$ for $1 \leq k \leq n \leq 12$

$k \backslash n$	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	2	6	24	120	720	5,040	40,320	362,880	3,628,800	39,916,800
2	.	1	3	11	50	274	1,764	13,068	109,584	1,026,576	10,628,640	120,543,840
3	.	.	1	6	35	225	1,624	13,132	118,124	1,172,700	12,753,576	150,917,976
4	.	.	.	1	10	85	735	6,769	67,284	723,680	8,409,500	105,258,076
5	.	.	.	.	1	15	175	1,960	22,449	269,325	3,416,930	45,995,730
6	.	.	.	.	.	1	21	322	4,536	63,273	902,055	13,339,535
7	.	.	.	.	.	.	1	28	546	9,450	157,773	2,637,558
8	.	.	.	.	.	.	.	1	36	870	18,150	357,423
9	.	.	.	.	.	.	.	.	1	45	1,320	32,670
10	.	.	.	.	.	.	.	.	.	1	55	1,925
11	.	.	.	.	.	.	.	.	.	.	1	66
12	.	.	.	.	.	.	.	.	.	.	.	1

Asymptotic properties, as $n \rightarrow \infty$ From equation (5) of Jordan, p. 160, we obtain, for fixed k ,

$$P(k, n) \sim \frac{[\log n]^{k-1}}{n(k-1)!}.$$

Thus $\lim_{n \rightarrow \infty} P(k, n) = 0$, uniformly in k . Let

$$P(k_{\max}, n) = \max_{k=1, \dots, n} P(k, n).$$

Then, for large n , we have approximately (Jordan, p. 161)

$$k_{\max} > \log n > k_{\max} - 1.$$

By (34)

$$\mu \sim \log n.$$

Hence

$$\bar{l} \sim \frac{n}{\log n}.$$

Thus both μ and $\bar{l} \rightarrow \infty$; however,

$$\frac{\bar{l}}{\mu} \sim \frac{n}{[\log n]^2} \rightarrow \infty.$$

By (35), $\sigma^2 \sim \log n$, also.

As we might expect, a central limit theorem applies, viz. prob. (no. of blocks yielded by a R.A. of $[n]$ lies between $\log n + a\sqrt{\log n}$ and $\log n + b\sqrt{\log n}$, $a < b$) $\sim \Phi(b) - \Phi(a)$. For a justification of this, see Feller (1957, pp. 242-3), where problem (a) of § 8 is discussed.

8. TWO RELATED PROBLEMS CONCERNING RANDOM PERMUTATIONS

(a) The following problem is treated in Riordan (1958, pp. 66-72) and in Feller (1957, pp. 242-3).

'What is the probability $P'(k, n)$ that a random permutation $\sigma \begin{pmatrix} 1 \dots n \\ \sigma_1 \dots \sigma_n \end{pmatrix}$ yields k cycles?'

Surprisingly, it turns out that we may derive the equations (10)-(13) for $P'(k, n)$ in a similar, but more elementary, way and so $P'(k, n) \equiv P(k, n)$.

(b) The second analogous problem may be stated thus:

'Let $\sigma(\sigma_1 \dots \sigma_n)$ be a random permutation of the integers $1, \dots, n$. Take as first "block" $(\sigma_1 \dots \sigma_{j_1})$, where $\sigma_{j_1} = \max_{j=1, \dots, n} \sigma_j (= n)$; $(\sigma_{j_1+1} \dots \sigma_{j_2})$ as second "block", where $\sigma_{j_2} = \max_{j=j_1+1, \dots, n} \sigma_j$; and so on. What is the probability $P''(k, n)$ that this amalgamation process yields k blocks in all?'

By a simple argument (in which the position of n is considered) we may obtain the recurrence relations

$$P''(1, n) = \frac{1}{n}$$

and

$$P''(k, n) = \frac{1}{n} \sum_{s=k-1}^{n-1} P''(k-1, s) \quad (2 \leq k \leq n).$$

It is easily verified that $H''(z, t) = H(z, t)$, where

$$H''(z, t) = \sum_{n=1}^{\infty} \sum_{k=1}^n P''(k, n) t^k z^n, \quad \text{and so} \quad P''(k, n) = P(k, n).$$

This result may alternatively be derived as a corollary of the above theory, as follows. Select an n -collection $[n]_{\mathcal{D}}$ of ordinates X_i , with $X_n > \dots > X_1$, and weights $1, \dots, 1$ which satisfy the additional condition:

$$\mathcal{D}: X_{\rho_1} + \dots + X_{\rho_r} > \quad \text{or} \quad < X_{\rho_1} + \dots + X_{\rho_{r+s}}$$

according as

$$\max_{i=1, \dots, r} \rho_i > \quad \text{or} \quad \leq \max_{i=1, \dots, (r+s)} \rho_i,$$

for any $(r+s)$ -permutation $\rho(\rho_1 \dots \rho_{r+s})$ of the integers $1, \dots, n$, where $1 \leq r, s \leq n-1$ and $2 \leq r+s \leq n$.

We may use a step-by-step construction to verify that such sets of ordinates exist. With the n weights $1, \dots, 1$ select arbitrarily the ordinates X_1 and X_2 , with $X_2 > X_1$; X_1 and X_2 satisfy $\mathcal{C}$ (with weights $1, 1$) and $\mathcal{D}$. Now choose $X_3 > X_2$, and sufficiently large that X_1, X_2 and X_3 satisfy $\mathcal{C}$ (with weights $1, 1, 1$) and $\mathcal{D}$. Likewise choose successively $X_4, X_5, \dots, X_n$.

Now consider more closely the complete amal. by $\mathcal{A}_2$ of the R.A. $\begin{pmatrix} X_{\sigma_1} \dots X_{\sigma_n} \\ 1 \dots 1 \end{pmatrix}$ of $[n]_{\mathcal{D}}$. The first block will be $Y_{j_1}^1$, where $X_{j_1}^1 = \max_{j=1, \dots, n} X_{\sigma_j}^1$. But, by $\mathcal{D}$, $X_{j_1} = \max_{j=1, \dots, n} X_{\sigma_j} (= X_n)$. Similarly, the second block will be $Y_{j_2}^{j_1+1}$, where $X_{j_2} = \max_{j=j_1+1, \dots, n} X_{\sigma_j}$; and so on.

It is now seen that, for the same random permutation σ , the amalgamation process in the two problems is identical, under the correspondence $\sigma_i \leftrightarrow X_{\sigma_i}$. Hence $P''(k, n) = P(k, n)$.

The problems (a) and (b) lie primarily in the realm of combinatorial analysis rather than probability theory (as perhaps also does the rest of the work).

I am indebted to both Dr D. J. Bartholomew, for originating the problem, and Mr D. V. Lindley, for bringing it to my notice and giving helpful advice.

REFERENCES

- BARTHOLOMEW, D. J. (1959a). A test of homogeneity for ordered alternatives. *Biometrika*, **46**, 36-48.
 BARTHOLOMEW, D. J. (1959b). A test of homogeneity for ordered alternatives. II. *Biometrika*, **46**, 328-35.
 FELLER, W. (1957). *An Introduction to Probability Theory and its Applications*, 2nd ed. New York: John Wiley and Sons.
 JORDAN, C. (1947). *Calculus of Finite Differences*, 2nd ed. New York: Chelsea Publishing Company.
 RIORDAN, J. (1958). *An Introduction to Combinatorial Analysis*. New York: John Wiley and Sons.

APPENDIX

We are now in a position to prove the truth of Bartholomew's conjecture (*)-(**) for $\hat{P}(l, k)$ referred to in § 1. First, applying the Corollary to Theorem 2 with the probability distribution taken to be normal, it is seen that $\hat{P}(l, k) = P(l, k)$, where $\hat{P}(l, k)$ is as defined in § 1. Thus

$$\hat{P}(l, k) = P(l, k) = \frac{1}{k!} |S_k^l|.$$

To complete the proof, we prove that the conjectured recurrence relations

$$\hat{P}(l, k) = \frac{1}{l} \sum_{j=l-1}^{k-1} \hat{P}(1, k-j) \hat{P}(l-1, j) \quad (2 \leq l \leq k), \quad (*)$$

$$\hat{P}(1, k) = \frac{1}{k}, \quad (**)$$

where

yield the same explicit formula for $\hat{P}(l, k)$ as above, hence thereby ensuring their own validity. Using (**), substitute for $\hat{P}(1, k-j)$ in (*). Multiply the resulting equation by z^k and sum from $k = 1$ to ∞ . Writing

$$\hat{F}_l(z) = \sum_{k=l}^{\infty} \hat{P}(l, k) z^k,$$

this gives

$$\begin{aligned} \hat{F}_l(z) &= \frac{1}{l} \left\{ z \hat{F}_{l-1}(z) + \frac{1}{2} z^2 \hat{F}_{l-1}(z) + \dots + \frac{1}{s} z^s \hat{F}_{l-1}(z) + \dots \right\} \\ &= \frac{1}{l} \log \left(\frac{1}{1-z} \right) \hat{F}_{l-1}(z). \\ \hat{F}_1(z) &= z + \frac{1}{2} z^2 + \dots + \frac{1}{s} z^s + \dots = \log \left(\frac{1}{1-z} \right). \end{aligned}$$

Also, by (**),

$$\text{From the last two equations, we have} \quad \hat{F}_l(z) = \frac{[\log \{1/(1-z)\}]^l}{l!} = F_l(z),$$

by (32). Hence $\hat{P}(l, k) = P(l, k)$ for all l, k . Thus, while not as simple as (10)-(13), (*)-(**) are indeed exactly equivalent to them. It does not seem possible to prove (*) directly despite the heuristic interpretation it may be given. However, (*) is but one of a number of recurrence relations satisfied by the $P(l, k)$. We have one in § 8(b) for $P''(l, k) \equiv P(l, k)$, and yet more may be found in Jordan (1947, Ch. IV), by using the relation $|S_k^l| = k! P(l, k)$.

Note added by author in proof. I am grateful to Dr H. D. Brunk of the University of Missouri for the information that this type of problem has been considered before by E. S. Andersen (*Math. Scand.* (1954), **2**, §§ 7, 8, pp. 209-18). He derives ((8.31), p. 216) my recurrence relations (10)-(13), using an entirely different, but less direct, method. Apart from this, there is little or no overlap between Andersen's paper and mine.

A TEST OF HOMOGENEITY FOR ORDERED ALTERNATIVES. II

BY D. J. BARTHOLOMEW

University College of North Staffordshire

1. INTRODUCTION AND SUMMARY

In an earlier paper (Bartholomew, 1959) referred to subsequently as I, modified χ^2 - and F -tests were introduced for use against ordered alternatives. The need for them arises in the following way. Suppose that we have a sample of k independent values $x_1, x_2, \dots, x_k$, where x_i is normally distributed with mean m_i and standard deviation σ_i ($i = 1, 2, \dots, k$). A test is required of the hypothesis $m_1 = m_2 = \dots = m_k$ (H_0) against the ordered alternative $m_1 \geq m_2 \geq \dots \geq m_k$ (H_2). The modified tests, denoted by $\bar{\chi}^2$ and $\bar{F}$, are calculated after the original values have been 'reduced' by an averaging process which depends upon their order and magnitude.

The distribution of these statistics was completely determined for $k = 3$ and $k = 4$ and tables of percentage points were given for $\bar{\chi}^2$. The distribution depends upon certain probabilities, $P(l, k)$. In the general case, when the σ 's are not equal, their determination for $k > 4$ was not possible because of lack of knowledge about the normal multivariate integral. For the special case of equal standard deviations, however, a recurrence relation was conjectured. Miles (1959) has proved this conjecture and has further shown that $P(l, k) = |S_k^l|/k!$, where S_k^l is the Stirling number of the first kind. This result has made it desirable to extend the tables of percentage points and to give some further properties of the distribution. This work occupies § 2 of the present paper.

It was pointed out in I that a test was required for use when the direction of the ordering was unknown; that is, for the alternative

$$H_3: m_1 \geq m_2 \geq \dots \geq m_k, \quad \text{or} \quad m_k \geq m_{k-1} \geq \dots \geq m_1.$$

Such a test is provided in § 3. It is a two-sided version of $\bar{\chi}^2$ (or $\bar{F}$) and requires no new distribution theory. Some extra significance levels have been given for $k = 3$ and $k = 4$ to facilitate the use of this version of the test.

A good approximation to the distribution in the general case, for $k = 5$, has been obtained; this is given in § 4.

2. THE DISTRIBUTION OF $\bar{\chi}^2$ FOR EQUAL WEIGHTS

2.1. The distribution of $\bar{\chi}^2$ when $\sigma_1 = \sigma_2 = \dots = \sigma_k$ is given by

$$\left. \begin{aligned} \Pr\{\bar{\chi}^2 \geq \gamma\} &= \frac{1}{k!} \sum_{l=2}^k |S_k^l| \int_{\gamma}^{\infty} p(\chi_{l-1}^2) d\chi_{l-1}^2, \\ \Pr\{\bar{\chi}^2 = 0\} &= 1/k. \end{aligned} \right\} \quad (1)$$

A table of Stirling numbers will be found in Miles's paper. The 10, 5, 2.5, 1 and 0.5% points of $\bar{\chi}^2$ calculated from (1) are given in Table 1.

Table 1. *Percentage points of $\bar{\chi}_k^2$ when the weights are equal*

$100P$ k	10	5	2.5	1	0.5
3	2.580	3.820	5.098	6.822	8.146
4	3.187	4.528	5.891	7.709	9.092
5	3.636	5.049	6.471	8.356	9.784
6	3.994	5.460	6.928	8.865	10.327
7	4.289	5.800	7.304	9.284	10.774
8	4.542	6.088	7.624	9.639	11.153
9	4.761	6.339	7.901	9.946	11.480
10	4.956	6.560	8.145	10.216	11.767
11	5.130	6.758	8.363	10.458	12.025
12	5.288	6.937	8.561	10.676	12.257

It is of interest to point out that the distribution (1) is also applicable for certain sets of unequal σ 's. It can be shown that the general distribution depends on the $(k-2)$ parameters

$$-\rho_{i,i+1} = \sqrt{\left\{ \frac{a_i a_{i+2}}{(a_i + a_{i+1})(a_{i+1} + a_{i+2})} \right\}} \quad (i = 1, 2, \dots, k-2),$$

where $a_i = 1/\sigma_i^2$. These parameters are all equal to one-half when $a_1 = a_2 = \dots = a_k$, but it is also possible to find other sets of a 's for which this is so. The reason for this is that, for any set of ρ 's, there are $(k-2)$ equations to determine $(k-1)$ ratios, a_i/a_{i+1} ($i = 1, 2, \dots, k-1$). The distribution would therefore be described more correctly as that for equal ρ 's. However, because of the much greater practical importance of the case $a_1 = a_2 = \dots = a_k$ the description used in I has been retained.

2.2. *Properties of the distribution of $\bar{\chi}^2$ for equal weights.* The moments and limiting properties of the distribution are most conveniently approached via the characteristic function

$$\begin{aligned} \phi_{\bar{\chi}^2}(t) &= \sum_{l=1}^k P(l, k) (1 - 2it)^{-\frac{1}{2}(l-1)} \\ &= \frac{(z+1)(z+2)\dots(z+k-1)}{k!}, \end{aligned}$$

where $z = (1 - 2it)^{-\frac{1}{2}}$ (see Miles (1959) equation (30)). Hence, for the cumulant generating function we have

$$\psi(t) = \log_e \phi(t) = \sum_{j=1}^{k-1} \log_e (j+z) - \log_e k!.$$

The first four cumulants are found to be

$$\left. \begin{aligned} \kappa_1 &= \sum_2^k j^{-1}, & \kappa_2 &= \sum_2^k \{3j^{-1} - j^{-2}\}, \\ \kappa_3 &= \sum_2^k \{15j^{-1} - 9j^{-2} + 2j^{-3}\}, \\ \kappa_4 &= \sum_2^k \{105j^{-1} - 87j^{-2} + 36j^{-3} - 6j^{-4}\}. \end{aligned} \right\} \quad (2)$$

Some numerical values of β_1 and β_2 are given in Table 2, from which it will be seen that the distribution is extremely skew even when k is quite large. In the event of having to test significance for large k a curve could be fitted by moments. A χ^2 -curve, fitted using the first two moments, gives a very good approximation for $k = 10$. As $k \rightarrow \infty$

$$\phi(t) \rightarrow \frac{(k-1)^z}{\Gamma(z+1)k} \sim \frac{k^{z-1}}{\Gamma(z+1)},$$

using the Euler-Gauss product limit formula for the Γ -function. The cumulants of the limiting form are thus given by

$$\kappa_r = 1.3.5 \dots (2r-1) \cdot \log k.$$

The distribution ultimately approaches normality, but only very slowly, as shown in Table 2.

Table 2. *Exact and limiting values of β_1 and β_2 for $\bar{\chi}^2$ (equal weights)*

k	10	20	50	100	10^3	10^6	10^{12}
β_1 { Exact Limiting	4.14 3.62	3.10 2.78	2.30 2.13	1.94 1.81	— 1.21	— 0.60	— 0.30
β_2 { Exact Limiting	8.88 8.07	7.39 6.89	6.28 5.98	5.75 5.53	— 4.69	— 3.84	— 3.42

2.3. The distribution of $\bar{F}$, when the weights are equal, is given by

$$\left. \begin{aligned} \Pr\{\bar{F} \geq \gamma\} &= \frac{1}{k!} \sum_{l=2}^k |S_k^l| \int_{\gamma}^{\infty} p(F_{l-1, N-l}) dF \\ &= \frac{1}{k!} \sum_{l=2}^k |S_k^l| I_{1-\gamma}[\tfrac{1}{2}(N-l), \tfrac{1}{2}(l-1)], \\ \Pr\{\bar{F} = 0\} &= 1/k, \\ z &= (l-1)\gamma/[N-l+(l-1)\gamma]. \end{aligned} \right\} \quad (3)$$

where

Tables of percentage points have not been calculated, nor has any simple approximation been found. However, if tables of the incomplete B -function are available, significance can be tested as described in I.

3. A TWO-SIDED TEST

3.1. In some problems it is possible to rank the means under the alternative without knowing whether the sequence is increasing or decreasing. This hypothesis may be written

$$H_3: m_1 \geq m_2 \geq \dots \geq m_k \quad \text{or} \quad m_k \geq m_{k-1} \geq \dots \geq m_1.$$

A test designed to detect this kind of departure will be called two-sided. The experiment to compare two ulcer treatments (A and B), discussed in I, would have required this kind of test had it not been assumed that treatment A would be better than B under the alternative.

A further example, to illustrate the method, is given below. The following test procedure is suggested. First, calculate $\bar{\chi}^2$ (or $\bar{F}$) as if the alternative were $m_1 \geq m_2 \geq \dots \geq m_k$, and then as if it were $m_k \geq m_{k-1} \geq \dots \geq m_1$. If the two values of $\bar{\chi}^2$ so obtained are denoted by $\bar{\chi}^{2'}$ and $\bar{\chi}^{2''}$, the test criterion is defined as $\max(\bar{\chi}^{2'}, \bar{\chi}^{2''})$. Since $\bar{\chi}^2$ can be derived by the likelihood ratio technique it follows that the two-sided version can also be obtained in this way.

Example

The table given below has been constructed from data given by Maguire, Pearson & Wynn (1952, 1953) on explosions in coal mines involving more than ten men killed. It illustrates the application of the two-sided test and draws attention to a further field of application. The null hypothesis to be tested is that the explosions occur randomly at a constant rate. The alternative is that the rate of occurrence has either increased or decreased during the period. If the alternative stated only that the rate differed from one period to another the Poisson index of dispersion would be used to test the null hypothesis. The application of $\bar{\chi}^2$ to the frequencies is equivalent to calculating this index for the reduced problem. It has to be assumed that the Poisson frequencies are approximately normally distributed; this assumption is implied in the use of the χ^2 -distribution to test the significance of the index of dispersion and will be considered adequate here. We therefore have to test the homogeneity of a set of normal variables with weights equal to the reciprocals of the mean frequency. After carrying out the averaging process indicated on the right of the table we arrive at the following two reduced forms.

Table 3. Application of the two-sided $\bar{\chi}^2$ -test to data on the number of explosions in coal mines involving more than ten men killed (1876-1945)

Period	No. of explosions	Decreasing trend	Increasing trend
1876-85	31		$\left. \begin{array}{l} 20 \\ 18.5 \end{array} \right\} 13.2$
1886-95	22		
1896-05	7	$\left. \begin{array}{l} 10.5 \\ 10.0 \end{array} \right\} 11.0$	
1906-15	14		$\left. \begin{array}{l} 13.2 \\ 14.5 \end{array} \right\}$
1916-25	5		
1926-35	15		
1936-45	14		

Mean frequency = 15.43.

(a) Decreasing trend. Average: $\begin{array}{ccc} 31 & 22 & 11 \\ & 1 & 5 \end{array}$
 Weight $\times 15.43$: $\begin{array}{ccc} & & \\ & 13.2 & 14.5 \end{array}$

(b) Increasing trend. Average: $\begin{array}{ccc} & & \\ & 5 & 2 \end{array}$
 Weight $\times 15.43$: $\begin{array}{ccc} & & \end{array}$

It is clear that $\bar{\chi}^2$ calculated from (a) will be much larger than from (b) so we have

$$\max(\bar{\chi}^{2'}, \bar{\chi}^{2''}) = \{31^2 + 22^2 + 5(11)^2 - 7(15.43)^2\} / 15.43$$

$$= 24.85.$$

It will be clear from the following section, and Table 1, that this value is highly significant and therefore that the data provide strong evidence for a decreasing trend. This conclusion is in agreement with those reached using other tests (Maguire *et al.* (1953), Barnard (1953)).

3.2. *The distribution of the two-sided test.* Let E_1 denote the event $\bar{\chi}^{2'} \geq \gamma$ and E_2 the event $\bar{\chi}^{2''} \geq \gamma$. The probability integral of the two-sided test is given by

$$P(E_1 + E_2) = P(E_1) + P(E_2) - P(E_1 E_2).$$

By symmetry, $P(E_1) = P(E_2)$ so that if γ is the $100\alpha\%$ point of $\bar{\chi}^2$.

$$P(E_1 + E_2) = 2\alpha - P(E_1 E_2).$$

It can be deduced immediately from this that

$$\alpha \leq P(E_1 + E_2) \leq 2\alpha.$$

In the case $k = 3$ it is possible to determine $P(E_1 E_2)$ exactly, using tables of the normal bivariate integral; these calculations show that it can be neglected by comparison with 2α . The following heuristic argument indicates that this will be true for all k .

If γ is large (so that α is small) the events E_1 and E_2 indicate significant departures from the null hypothesis in opposite directions. Consequently, the fact that one has occurred reduces the chance that the other will also have occurred, that is $P(E_2 | E_1) < \alpha$. Thus we have

$$P(E_1 E_2) = P(E_1) P(E_2 | E_1) < \alpha^2.$$

Since α is small, we take

$$\Pr\{\max(\bar{\chi}^{2'}, \bar{\chi}^{2'') \geq \gamma\} \doteq 2 \Pr\{\bar{\chi}^2 \geq \gamma\}$$

when testing significance, the maximum possible error being α^2 . The existing tables can therefore be used if the significance level is doubled. The value obtained for the example in § 3.1 for $k = 7$ lies well beyond the 1% point for the two-sided test.

3.3. In order to provide 5 and 1% points for the two-sided test, 2.5 and 0.5% levels of $\bar{\chi}^2$ have been computed. They are given in tables 4 and 5 for $k = 3$ and $k = 4$. For convenience, Table 4 also includes the 5 and 1% levels already given in I and an additional column of 10% points.

Table 4. Percentage points of $\bar{\chi}_3^2$

$\begin{matrix} 100P \\ \diagdown \\ -\rho_{12} \end{matrix}$	10	5	2.5	1	0.5
0.0	2.952	4.231	5.537	7.289	8.628
0.1	2.885	4.158	5.459	7.208	8.543
0.2	2.816	4.081	5.378	7.122	8.455
0.3	2.742	4.001	5.292	7.030	8.360
0.4	2.664	3.914	5.200	6.932	8.258
0.5	2.580	3.820	5.098	6.822	8.146
0.6	2.486	3.715	4.985	6.700	8.016
0.7	2.379	3.593	4.852	6.556	7.865
0.8	2.251	3.446	4.689	6.377	7.677
0.9	2.080	3.245	4.465	6.130	7.413
1.0	1.642	2.706	3.841	5.413	6.635

Table 5. The 2.5 % (upper figure) and 0.5 % (lower figure) points of $\bar{\chi}_4^2$

$-\rho_{12}$ $-\rho_{23}$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.0	6.861 10.171	— —	— —	— —	— —	— —	— —	— —
0.1	6.795 10.098	6.724 10.021	— —	— —	— —	— —	— —	— —
0.2	6.725 10.022	6.649 9.939	6.570 9.851	— —	— —	— —	— —	— —
0.3	6.653 9.942	6.571 9.852	6.484 9.756	6.391 9.653	— —	— —	— —	— —
0.4	6.575 9.855	6.486 9.758	6.392 9.653	6.289 9.540	6.174 9.411	— —	— —	— —
0.5	6.491 9.761	6.394 9.654	6.289 9.538	6.174 9.410	6.043 9.264	5.891 5.702	— 5.462	— —
0.6	6.398 9.656	6.289 9.536	6.171 9.405	6.038 9.257	5.886 9.086	8.877 5.443	8.604 5.100	— 4.346
0.7	6.292 9.534	6.166 9.396	6.028 9.242	5.870 9.065	5.682 8.853	8.581 4.999	8.183 3.841	7.278 —
0.8	6.163 9.385	6.011 9.217	5.838 9.025	5.634 8.795	5.375 8.499	— —	— —	— —
0.9	5.991 9.183	5.778 8.948	5.523 8.661	5.183 8.273	4.591 7.577	— —	— —	— —
1.0	5.537 8.628	— —	— —	— —	— —	— —	— —	— —

4. THE DISTRIBUTION FOR $k = 5$

4.1. Unless all the correlations are equal to minus one-half it is not possible to test significance exactly for large k . Existing results suggest that the distribution for equal weights may be a good approximation unless there is a large divergence of the ρ 's from minus one-half. Alternatively, the number of groups can be reduced by pooling as suggested in I. Fortunately, a great many practical problems fall within the scope of the existing theory. However, gradings of the kind—very good, good, fair, poor and bad—or—extreme, moderate and uncommitted—occasionally run to as many as five or six categories. It is therefore desirable to obtain some results for these cases. In what follows we give a good approximation for the case $k = 5$ and suggest that the same method might yield useful results for $k = 6$. The only difficulty in the determination of the exact distribution lies in finding $P(5, 5)$. Plackett (1954) gives a method of obtaining exact values which involves numerical integration. However, this arithmetic can be avoided by using an approximate formula given by the same author. $P(5, 5)$ is the volume in the positive quadrant of the four-variate normal distribution with correlation matrix

$$R = \begin{bmatrix} 1 & \rho_{12} & 0 & 0 \\ \rho_{12} & 1 & \rho_{23} & 0 \\ 0 & \rho_{23} & 1 & \rho_{34} \\ 0 & 0 & \rho_{34} & 1 \end{bmatrix}.$$

Plackett gives the approximate formula

$$P(5, 5) = \frac{1}{4\pi^2} [\cos^{-1}(-\rho_{12}) \cos^{-1}(-\rho_{34}) + \frac{1}{2}\pi \cos^{-1}(-\rho_{23})] - \frac{1}{16}$$

valid for small ρ . It is shown below that this approximation is very good for the present purpose when $-\rho_{12} = -\rho_{23} = -\rho_{34} = \frac{1}{2}$, and cannot lead to a large error for larger ρ 's.

4.2. Using Plackett's approximate formula for $P(5, 5)$ and the same method as in I, we obtain the following probabilities. They are expressed in terms of $\sin^{-1}$ since this function has been extensively tabled in the *Table of Arcsin x* (1945). The formulae have been simplified by using the notation of partial correlation.

$$P(5, 5) = \{\sin^{-1}(-\rho_{12}) \sin^{-1}(-\rho_{34})\} / (4\pi^2)$$

$$P(4, 5) = \frac{1}{4} - \{\sin^{-1}(-\rho_{12}) + \sin^{-1}(-\rho_{23}) + \sin^{-1}(-\rho_{34})\} / (8\pi) + \frac{1}{16},$$

$$- \{\sin^{-1}(-\rho_{12,3}) + \sin^{-1}(-\rho_{23,1}) + \sin^{-1}(-\rho_{34,2}) + \sin^{-1}(-\rho_{13,2}) + \sin^{-1}(-\rho_{23,4})$$

$$+ \sin^{-1}(-\rho_{24,3})\} / (8\pi),$$

$$P(3, 5) = \frac{1}{2} - \{\sin^{-1}(-\rho_{12,34}) + \sin^{-1}(-\rho_{13,24}) + \sin^{-1}(-\rho_{14,23}) + \sin^{-1}(-\rho_{23,14})$$

$$+ \sin^{-1}(-\rho_{24,13}) + \sin^{-1}(-\rho_{34,12})\} / (8\pi)$$

$$+ \{\sin^{-1}(-\rho_{12}) + \sin^{-1}(-\rho_{23}) + \sin^{-1}(-\rho_{34})\} / (8\pi)$$

$$- \{\sin^{-1}(-\rho_{12}) \sin^{-1}(-\rho_{34,12}) + \sin^{-1}(-\rho_{23}) \sin^{-1}(-\rho_{14,23})$$

$$P(2, 5) = \frac{1}{2} - P(4, 5),$$

$$+ \sin^{-1}(-\rho_{34}) \sin^{-1}(-\rho_{12,34})\} / (4\pi^2),$$

$$P(1, 5) = 1 - \sum_{j=2}^5 P(j, 5).$$

4.3. A comparison between the exact and approximate probability integral can be made for the case $-\rho_{12} = -\rho_{23} = -\rho_{34} = \frac{1}{2}$. In this case the exact value of $P(5, 5)$ is $1/120$ and the approximate value $1/144$. This discrepancy has a negligible effect on the significance level as can be seen from Table 6. In fact even if $P(5, 5)$ is put equal to zero the effect is not serious.

Table 6. Comparison of the probability integral given by the exact and approximate formulae for $P(5, 5)$ when $-\rho_{12} = -\rho_{23} = -\rho_{34} = \frac{1}{2}$

$\gamma \backslash P$	Exact	Plackett's approximation	$P(5, 5) = 0$
5.049	0.05000	0.04961	0.04765
8.356	0.01000	0.00989	0.00933

Plackett also gives a table of exact values of $P(5, 5)$ for various combinations of large ρ 's. In all cases $P(5, 5)$ is very small so that any error involved in using the approximate formula will be unimportant. In some cases, especially when the approximate method yields a negative value, it would obviously be preferable to put $P(5, 5) = 0$. Tabulation of percentage points would require a table with three-way entry, although the size could be reduced by making use of symmetry and the exclusion of extreme values of the ρ 's.

The method used here would give a partial solution for $k > 5$. For example, when $k = 6$ we could assume that $P(6, 6)$ was negligible, obtain an approximate value for $P(5, 6)$ and determine the remaining probabilities exactly. This possibility has not been explored.

REFERENCES

- BARNARD, G. A. (1953). Time intervals between accidents—a note on Maguire, Pearson and Wynn's paper. *Biometrika*, **40**, 212–13.
- BARTHOLOMEW, D. J. (1959). A test of homogeneity for ordered alternatives. *Biometrika*, **46**, 36–48.
- MAGUIRE, B. A., PEARSON, E. S. & WYNN, A. H. A. (1952). The time intervals between industrial accidents. *Biometrika*, **39**, 168–80.
- MAGUIRE, B. A., PEARSON, E. S. & WYNN, A. H. A. (1953). Further notes on the analysis of accident data. *Biometrika*, **40**, 213–16.
- MILES, R. E. (1959). The complete amalgamation into blocks, by weighted means, of a finite set of real numbers. *Biometrika*, **46**, 317–27.
- PLACKETT, R. L. (1954). A reduction formula for normal multivariate integrals. *Biometrika*, **41**, 351–60.
- Table of Arcsin x (1945). Mathematical Tables Project No 17, sponsored by the National Bureau of Standards. New York: Columbia University Press.

THE MULTIPLE-RECAPTURE CENSUS

II. ESTIMATION WHEN THERE IS IMMIGRATION OR DEATH

By J. N. DARROCH

Statistical Laboratory, University of Manchester

1. INTRODUCTION

1.1. In a previous paper (1958), which will be referred to as (I), we discussed the multiple-recapture census when the population is closed both to augmentation from outside and departure from inside. These restrictions are now removed.

Let the experimenter take s samples, as in (I). Also, let $1 - \phi_k$ be the conditional probability that an individual dies (or permanently emigrates) between the k th and $(k + 1)$ th samples given that it is alive at the time of the k th, $k = 1, 2, \dots, s - 1$. Let n_1 be the size of the population at the time of the first sample and let $n_k - n_{k-1}$ new individuals immigrate (or be born) into the population between the $(k - 1)$ th and k th samples and be alive at the time of the k th, $k = 2, 3, \dots, s$. Nothing is assumed about the $n_k - n_{k-1}$ and they are treated as parameters of the model. To treat them as random variables would entail assumptions about the manner in which they vary and would complicate rather than simplify the probability densities.

In § 2 we shall take $\phi_k = 1$ in which case n_i is the population size at the time of the i th sample, $i = 1, 2, \dots, s$. In § 3, $n_i = n$ and the size of the population at the time of the i th sample ($i > 1$) is a random variable with expected value $n\phi_1 \dots \phi_{i-1}$. The general case when there is both immigration and death will be considered in § 4.

1.2. The main aims of this paper, as of (I), are to provide exact, fully stochastic models for the observed frequencies of individuals, to show how simply these frequencies naturally group themselves, and to obtain estimates of the unknown parameters. When there is immigration only or death only, the estimates are shown to be asymptotically efficient and their variances are found. In addition, a method of performing tests on the values of the parameters is given. When both immigration and death are operating, on the other hand, the complexity of the probability density prevents us from going further than obtaining the estimates and merely indicating how their variances can be found.

Both in (I) and in the present work we have been unable to obtain satisfactory tests of the underlying assumptions of the models, notably that tagged and untagged animals are captured with equal probabilities. (The above-mentioned tests on the values of the parameters assume the truth of the models.) We hope to fill this gap at a later date.

1.3. We gave a brief review of the literature and a list of references in (I) but, unfortunately, omitted to refer to a paper by Leslie, Chitty & Chitty (1953), the third of three papers by these authors. These three papers contain several ingenious mathematical approaches to multiple-recapture problems, together with a very full discussion of field data on populations of voles, and are uniquely valuable in the way that they dove-tail theory with practice. One of the points emphasized by the authors is this: although a basic feature of the multiple-recapture census is that caught animals are returned to the population alive and unharmed, in practice a few are either accidentally killed or have to be removed from the population.

This eventuality has not been taken account of in the following pages but we note here that it is easy to do so. Each of the usual observed classes 'caught in the $i, \dots, l$ samples but not otherwise' is coupled with an extra one for those animals which are caught in the same samples but killed or removed at the l th. Also, an additional parameter is introduced to represent the probability that a captured animal is killed or removed at any sample. The estimates of the other parameters change slightly but, apart from this, there are no complications.

2. IMMIGRATION BUT NO DEATH

2.1. In this section, suffices i, j take all values from 1 to s while suffices k, l take all values from 2 to s .

Let $p_i (= 1 - q_i)$ denote the probability that any member of the population is caught at the i th sample and let a_i be the size of the i th sample. Let $a_{<k}$ be the number of individuals caught before the k th sample and $a_{<k,k}$ the number which are caught before and at the k th sample. Further, let w denote any non-empty subset of the integers 1, 2, ..., s and u_w the number caught in every one of the w -samples but not otherwise. Let

$$r = \sum_w u_w = \sum_i u_i + \sum_{i < j} u_{ij} + \dots + u_{12\dots s},$$

the total number of individuals caught in the whole experiment.

The notation thus far is the same as in (I) and suffices for most of § 2. However, for the purpose of deriving the density $p[\{u_w\}]$, it is convenient to have a further notation. Namely, let a_w be the number of individuals caught in every one of the w -samples, regardless of whether or not they are caught otherwise.

We begin by finding $p[\{u_w\}]$ for $s = 3$, using a chainwise argument. The probability of catching a_1 individuals in the first sample is

$$\binom{n_1}{a_1} p_1^{a_1} q_1^{n_1 - a_1}.$$

Given a_1 , the probability of catching a_2 in the second, of which a_{12} are common to the first and $a_2 - a_{12}$ are new ones, is

$$\binom{a_1}{a_{12}} \binom{n_2 - a_1}{a_2 - a_{12}} p_2^{a_2} q_2^{n_2 - a_2}.$$

Given a_1, a_2, a_{12} , the probability of catching a_3 in the third, of which a_{123} are common to the first and second, $a_{13} - a_{123}$ to the first only, $a_{23} - a_{123}$ to the second only and of which $a_3 - a_{13} - a_{23} + a_{123}$ are new ones is

$$\binom{a_{12}}{a_{123}} \binom{a_1 - a_{12}}{a_{13} - a_{123}} \binom{a_2 - a_{12}}{a_{23} - a_{123}} \binom{n_3 - a_1 - a_2 + a_{12}}{a_3 - a_{13} - a_{23} + a_{123}} p_3^{a_3} q_3^{n_3 - a_3}.$$

Multiplying these three expressions together, cancelling and renaming the terms involved, we obtain

$$p[\{u_w\}] = \frac{n_1!}{(n_1 - a_{<2})!} \frac{(n_2 - a_{<2})!}{(n_2 - a_{<3})!} \frac{(n_3 - a_{<3})!}{(n_3 - r)!} \frac{1}{\prod_w u_w!} \prod_{i=1}^3 p_i^{a_i} q_i^{n_i - a_i}. \quad (1)$$

Note that, if $n_i = n$, (1) contracts to

$$\frac{n!}{(n - r)! \prod_w u_w!} \prod_i p_i^{a_i} q_i^{n - a_i},$$

which is the model forming the basis of (I).

(1) may also be written

$$p[\{u_w\}] = \frac{n_1!}{(n_1 - a_{<2})! u_1! u_{12}! u_{13}! u_{123}!} q_1^{n_1 - a_{<2}} (p_1 q_2 q_3)^{u_1} (p_1 p_2 q_3)^{u_{12}} (p_1 q_2 p_3)^{u_{13}} (p_1 p_2 p_3)^{u_{123}} \\ \times \frac{(n_2 - a_{<3})!}{(n_2 - a_{<3})! u_2! u_{23}!} q_2^{n_2 - a_{<3}} (p_2 q_3)^{u_2} (p_2 p_3)^{u_{23}} \\ \times \frac{(n_3 - a_{<3})!}{(n_3 - r)! u_3!} q_3^{n_3 - r} p_3^{u_3}. \quad (1')$$

Another way of exhibiting the distribution of the u_w is by means of their generating function $\psi[\{t_w\}] = E[\prod_w t_w^{u_w}]$. It is soon shown that

$$\psi[\{t_w\}] = (q_1 q_2 q_3 + p_1 q_2 q_3 t_1 + p_1 p_2 q_3 t_{12} + p_1 q_2 p_3 t_{13} + p_1 p_2 p_3 t_{123} \\ + q_1 p_2 q_3 t_2 + q_1 p_2 p_3 t_{23} + q_1 q_2 p_3 t_3)^{n_1} \\ \times (q_2 q_3 + p_2 q_3 t_2 + p_2 p_3 t_{23} + q_2 p_3 t_3)^{n_2 - n_1} \\ \times (q_3 + p_3 t_3)^{n_3 - n_2}. \quad (2)$$

2.2. Let us define $a_{<1} = 0$ and $a_{<s+1} = r$. Then for general s , (1) becomes

$$p[\{u_w\}] = \frac{1}{\prod_w u_w!} \prod_i \frac{(n_i - a_{<i})!}{(n_i - a_{<i+1})!} p_i^{a_i} q_i^{n_i - a_i}. \quad (3)$$

Maximizing with respect to p_i by equating $(\partial/\partial p_i) \log p[\{u_w\}]$ to zero,

$$\tilde{n}_i \tilde{p}_i = a_i. \quad (4)$$

Maximizing with respect to n_i by equating $\Delta_{n_i} \log p[\{u_w\}]$ to zero,

$$\frac{\tilde{n}_i - a_{<i+1}}{\tilde{n}_i - a_{<i}} = \tilde{q}_i. \quad (5)$$

The equation for $i = 1$ in (4) is the same as that in (5) and there is consequently no information on n_1 and p_1 separately. Otherwise, combining (4) and (5) and remembering that $a_{<k+1} = a_{<k} + a_k - a_{<k,k}$, we obtain

$$\tilde{p}_k = \frac{a_{<k,k}}{a_{<k}} \quad (6)$$

and

$$\tilde{n}_k = \frac{a_{<k} a_k}{a_{<k,k}}$$

provided $a_{<k,k} > 0$.

2.3. Having found maximum likelihood estimates from the density (3) we cannot apply maximum likelihood large sample theory to them directly, for reasons given in (I) § 3.2. The principal obstacle is the presence of the n_i in the factorial terms of (3). To get over this difficulty, we use a suitably modified version of the argument of (I) § 5.7. That is, we apply maximum likelihood large sample theory to a density which is conditional on some of the sample variables and afterwards take expectations over these variables. In this case, the most suitable conditional density is $p[\{u_w\} | \{a_{<i+1}\}]$.

The conditional distribution of $a_{<i+1} - a_{<i}$ given $a_{<i}$ is $B[n_i - a_{<i}, p_i]$ (including $i = 1$ when the distribution of $a_{<2}$ is $B[n_1, p_1]$) and therefore

$$p[\{a_{<i+1}\}] = \prod_i \binom{n_i - a_{<i}}{a_{<i+1} - a_{<i}} p_i^{a_{<i+1} - a_{<i}} q_i^{n_i - a_{<i+1}}.$$

Hence

$$p[\{u_w\} | \{a_{<i+1}\}] = \frac{1}{\prod_w u_w!} \prod_i (a_{<i+1} - a_{<i})! \prod_k p_k^{a_{<k,k}} q_k^{a_{<k} - a_{<k,k}}. \quad (7)$$

Reverting for the moment to the parent density, we see that

$$\tilde{n}_k = (a_{<k+1} - a_{<k})/\tilde{p}_k + a_{<k}, \quad (8)$$

where $\tilde{p}_k$ is given by (6). This equation will prove useful for two reasons. First, $a_{<k}$ and $a_{<k+1}$ are held constant in the conditional density (7), and secondly, $\tilde{p}_k$ is the maximum likelihood estimate of p_k for the conditional as well as the parent density. In view of (8), it turns out to be preferable to change from p_k to $\theta_k = p_k^{-1}$. (8) then becomes

$$\tilde{n}_k = (a_{<k+1} - a_{<k})\tilde{\theta}_k + a_{<k}. \quad (9)$$

Let $L = L[\{\theta_k\}] = \log p[\{u_w\} | \{a_{<i+1}\}]$. Then

$$\frac{\partial L}{\partial \theta_k} = -\frac{a_{<k}}{\theta_k} + \frac{a_{<k} - a_{<k.k}}{\theta_k - 1},$$

which, when equated to zero, gives $\tilde{\theta}_k = a_{<k}/a_{<k.k}$ provided $a_{<k.k} > 0$. Further

$$\begin{aligned} -E\left[\frac{\partial^2 L}{\partial \theta_k^2} | \{a_{<i+1}\}\right] &= \frac{a_{<k}}{\theta_k^2(\theta_k - 1)}, \\ -E\left[\frac{\partial^2 L}{\partial \theta_k \partial \theta_l} | \{a_{<i+1}\}\right] &= 0 \quad (k \neq l), \end{aligned}$$

yielding the asymptotic formulae

$$E[(\tilde{\theta}_k - \theta_k)^2 | \{a_{<i+1}\}] = \frac{\theta_k^2(\theta_k - 1)}{a_{<k}},$$

$$E[(\tilde{\theta}_k - \theta_k)(\tilde{\theta}_l - \theta_l) | \{a_{<i+1}\}] = 0.$$

When $a_{<k.k} = 0$, the Taylor expansion of $\partial L / \partial \theta_k$ at $\tilde{\theta}_k$ about its value at θ_k is meaningless since there is no finite solution for $\tilde{\theta}_k$. However, the resulting invalidity of the last formulae is easily avoided by changing from $\tilde{\theta}_k$ to

$$\theta'_k = \frac{a_{<k} + 1}{a_{<k.k} + 1},$$

since θ'_k is always finite and has an asymptotic variance differing only negligibly from that of $\tilde{\theta}_k$ over the finite part of the latter's range. (This can be shown using the δ -technique.) Moreover, θ'_k has a negligible bias. For, given $\{a_{<i+1}\}$, $a_{<k.k}$ is $B[a_{<k}, p_k]$. Therefore

$$E[\theta'_k | \{a_{<i+1}\}] = E[\theta'_k | a_{<k}] = \theta_k(1 - q_k^{a_{<k}+1})$$

and the difference between this expression and θ_k may be neglected. Two further properties of the θ'_k should be mentioned here. First, they are independently distributed given $\{a_{<i+1}\}$. This follows from the independence of the variables $a_{<k.k}$, a property which is proved without difficulty. Secondly, using the δ -technique, we find that

$$\begin{aligned} E[(\theta'_k - \theta_k)^2 | \{a_{<i+1}\}] &= E[(\theta'_k - \theta_k)^2 | a_{<k}] \\ &= \frac{\theta_k^2(\theta_k - 1)}{a_{<k}} + O\left(\frac{1}{a_{<k}^2}\right), \end{aligned}$$

as $a_{<k} \rightarrow \infty$. Returning now to the estimation of the n_k and recalling (9), we replace $\tilde{n}_k$ by

$$n'_k = (a_{<k+1} - a_{<k})\theta'_k + a_{<k}, \quad (10)$$

$$n'_k = \frac{(a_{<k} + 1)(a_k + 1)}{a_{<k.k} + 1} - 1. \quad (11)$$

that is, by

Now

$$n'_k - n_k = (a_{<k+1} - a_{<k})(\theta'_k - \theta_k) + \theta_k[(a_{<k+1} - a_{<k}) - (n_k - a_{<k})p_k].$$

Therefore, taking expectations conditional on $\{a_{<i+1}\}$,

$$E[(n'_k - n_k) | \{a_{<i+1}\}] = -(a_{<k+1} - a_{<k}) \theta_k q_k^{a_{<k+1}} + \theta_k [(a_{<k+1} - a_{<k}) - (n_k - a_{<k}) p_k].$$

Now taking expectations over $a_{<2}, \dots, a_{<k-1}, a_{<k+1}, \dots, a_{<s+1}$,

$$\begin{aligned} E[(n'_k - n_k) | a_{<k}] &= -(n_k - a_{<k}) p_k \theta_k q_k^{a_{<k+1}} \\ &= -(n_k - a_{<k}) q_k^{a_{<k+1}}, \end{aligned}$$

since $a_{<k+1} - a_{<k} = B[n_k - a_{<k}, p_k]$. We shall neglect this bias.

Squaring (11) and taking expectations,

$$\begin{aligned} E[(n'_k - n_k)^2 | a_{<k}] &= [(n_k - a_{<k})^2 p_k^2 + (n_k - a_{<k}) p_k q_k] E[(\theta'_k - \theta_k)^2 | a_{<k}] \\ &\quad + 2\theta_k (n_k - a_{<k}) p_k q_k E[(\theta'_k - \theta_k) | a_{<k}] \\ &\quad + \theta_k^2 (n_k - a_{<k}) p_k q_k. \end{aligned} \quad (12)$$

In deriving an asymptotic formula for $E[(n'_k - n_k)^2]$, a convenient limit process to use is: $n_i \rightarrow \infty$ such that $n_i/n_j \rightarrow c_{ij}$, constant. To evaluate the expectation of the first term of the right-hand side of (12) over $a_{<k}$, consider separately the ranges $0 \leq a_{<k} \leq hn_1$ and $hn_1 < a_{<k} \leq n_{k-1}$, where $h < 1 - q_1 \dots q_{k-1}$ and hn_1 is integral. Now $a_{<k}$ is the sum of independent binomial variables $B[n_1, 1 - q_1 \dots q_{k-1}]$, $\dots$, $B[n_{k-1} - n_{k-2}, p_{k-1}]$ and therefore

$$P[a_{<k} \leq hn_1] \leq P[B[n_1, 1 - q_1 \dots q_{k-1}] \leq hn_1] = O\left(\frac{1}{\sqrt{n_1}} c^{n_1}\right),$$

where

$$c = \left(\frac{1 - q_1 \dots q_{k-1}}{h}\right)^h \left(\frac{q_1 \dots q_{k-1}}{1 - h}\right)^{1-h} < 1.$$

Also θ'_k can never exceed $a_{<k} + 1$. Therefore, for the range $0 \leq a_{<k} \leq hn_1$, the contribution of the first term of (12) is at most $O(n_k^2 n_1^{\frac{1}{2}} c^{n_1}) = O(n_1^{\frac{3}{2}} c^{n_1}) = o(1)$. For the other part of the range, we may use the asymptotic formula for the variance of θ'_k . Since the second term on the right-hand side of (12) is $o(1)$, we may thus far say that

$$\begin{aligned} E[(n'_k - n_k)^2] &= E_{a_{<k} > hn_1} \left[((n_k - a_{<k})^2 p_k^2 + (n_k - a_{<k}) p_k q_k) \right. \\ &\quad \left. \left(\frac{\theta_k^2 (\theta_k - 1)}{a_{<k}} + O\left(\frac{1}{a_{<k}^2}\right) \right) \right] + E_{a_{<k} \geq 0} [\theta_k^2 (n_k - a_{<k}) p_k q_k] + o(1). \end{aligned}$$

Let

$$\alpha_{<k} = E[a_{<k}] = n_1(1 - q_1 \dots q_{k-1}) + \dots + (n_{k-1} - n_{k-2}) p_{k-1}.$$

Then, expanding $1/a_{<k}$ about $1/\alpha_{<k}$ and taking expectations over $a_{<k}$, it will be found that

$$E[(n'_k - n_k)^2] = \frac{n_k(n_k - \alpha_{<k}) q_k}{\alpha_{<k} p_k} + O(1). \quad (13)$$

Note that, when all $n_i = n$, (13) becomes

$$E[(n'_k - n)^2] = \frac{n q_1 \dots q_k}{(1 - q_1 \dots q_{k-1}) p_k} + O(1),$$

which was derived previously in (I).

Thus, the n'_k are almost unbiased estimates of the n_k with variances given by (13) and, as is easily shown, with negligible covariances. Further, n'_k is the asymptotically most efficient estimate of the class of estimates

$$n_k^* = (a_{<k+1} - a_{<k}) \theta_k^* + a_{<k}$$

for which

$$E[(\theta_k^* - \theta_k)^2 | a_{<k}, a_{<k+1}] = \frac{\lambda_k}{a_{<k}} + O\left(\frac{1}{a_{<k}^2}\right)$$

and for which $E[(\theta_k^* - \theta_k) | a_{<k}, a_{<k+1}]$ is negligible. This is a simple consequence of the fact that, since θ_k' is asymptotically efficient for θ_k , $\theta_k^2(\theta_k - 1)$ is the minimum attainable value of λ_k . In (I) we pointed out that n_k' may be evaluated from 'similar' tagging and that, if $n_i = n$, the n_k' contain between them all of the information on n .

2.4. Suppose that catchability is constant throughout the experiment and that e_i units of effort are expended in catching the i th sample. Then

$$q_i = e^{-e_i/\gamma}, \quad (14)$$

where the unknown coefficient of e_i is written as a reciprocal parameter for much the same reason that, in § 2.3, we changed from p_k to $\theta_k = p_k^{-1}$. In previous effort models, it has been customary to make the expected catch size proportional to the effort. This is equivalent here to writing $p_i = e_i/\gamma$ which involves a relative error $O(e_i/\gamma)$, and the equivalent approximation for q_i is $q_i = 1$ which is obviously too severe. Therefore, accepting $q_i = 1 - e_i/\gamma$ as a satisfactory approximation to (14), that is, neglecting terms which are relatively $O(e_i^2/\gamma^2)$, the approximation of the same order for p_i is $p_i = e_i/\gamma(1 - e_i/2\gamma)$. When it is expedient to do so, we shall use these expressions. The fact that they lead to $p_i + q_i = 1 - e_i^2/2\gamma^2$ is not, of course, inconsistent with our decision to neglect terms $O(e_i^2/\gamma^2)$.

Again, consider the conditional density $p[\{u_w\} | \{a_{<j+1}\}]$ which gives rise to the likelihood

$$e^{L(\gamma)} = \prod_k (1 - e^{-e_k/\gamma})^{a_{<k,k}} (e^{-e_k/\gamma})^{a_{<k,k} - a_{<k,k}}.$$

Hence

$$\frac{\partial L}{\partial \gamma} = -\frac{1}{\gamma^2} \left[\sum_k \frac{a_{<k,k} e_k}{1 - e^{-e_k/\gamma}} - \sum_k a_{<k,k} e_k \right].$$

Equating this to zero, the equation for $\tilde{\gamma}$ is

$$\sum_k \frac{a_{<k,k} e_k}{(e_k/\tilde{\gamma})(1 - e_k/2\tilde{\gamma})} = \sum_k a_{<k,k} e_k$$

and replacing $(1 - e_k/2\tilde{\gamma})^{-1}$ by $1 + e_k/2\tilde{\gamma}$, we find that

$$\tilde{\gamma} = \frac{\sum_k (a_{<k,k} - \frac{1}{2}a_{<k,k}) e_k}{\sum_k a_{<k,k}}.$$

This formula breaks down if $\sum_k a_{<k,k} = 0$ (and so does the whole experiment since it has not yielded a single recapture). Therefore, instead of $\tilde{\gamma}$, consider

$$\gamma' = \frac{\sum_k (a_{<k,k} - \frac{1}{2}a_{<k,k}) e_k + \frac{1}{2}\bar{e}}{\sum_k a_{<k,k} + 1},$$

where $\bar{e} = \sum_k e_k/(s-1)$. γ' always exists and we suspect that it is almost unbiased, for the following reason. If all $e_k = \bar{e}$ and $p_k = \bar{p}$ say,

$$\gamma' = \bar{e} \frac{\sum_k a_{<k,k} + 1}{\sum_k a_{<k,k} + 1 - \frac{1}{2}\bar{e}}$$

and $\sum a_{<k,k} = \sum B[a_{<k,k}, \bar{p}] = B[\sum a_{<k,k}, \bar{p}]$ since the $a_{<k,k}$ are independent. Hence

$$E[\gamma' | \{a_{<j+1}\}] = \bar{e} \frac{1}{\bar{p}} [1 - \bar{q}^{\sum a_{<k,k} + 1}] - \frac{1}{2}\bar{e}.$$

Neglecting $\bar{q}^{\Sigma a_{<k+1}}$ and replacing $\bar{p}$ by $(\bar{e}/\gamma)(1 - \bar{e}/2\gamma)$,

$$E[\gamma' | \{a_{<j+1}\}] = \gamma.$$

In the general case when the e_k are unequal, it still seems reasonable to assume that γ' is almost unbiased.

To obtain $E[(\gamma' - \gamma)^2 | \{a_{<j+1}\}]$ we may invert

$$\begin{aligned} -E\left[\frac{\partial^2 L}{\partial \gamma^2} | \{a_{<j+1}\}\right] &= \gamma^{-4} \sum_k a_{<k} e_k^2 q_k / p_k \\ &= \gamma^{-2} \sum_k a_{<k} p_k \end{aligned}$$

on making use of our approximations for p_k and q_k . The δ -technique shows that

$$E[(\gamma' - \gamma)^2 | \{a_{<j+1}\}] = \frac{\gamma^2}{\sum a_{<k} p_k} + O((\sum a_{<k} p_k)^{-2}).$$

From now on, the argument follows very much the same lines as in § 2.3. For the parent density

$$\tilde{n}_i = (a_{<i+1} - a_{<i}) / \tilde{p}_i + a_{<i}$$

and $\tilde{\gamma}$ has the same value as for the conditional density. We replace it by γ' and $\tilde{n}_i$ by m'_i say, defined as

$$m'_i = (a_{<i+1} - a_{<i}) / (e_i / \gamma') (1 - e_i / 2\gamma') + a_{<i}.$$

Hence

$$m'_i - n_i = \frac{(a_{<i+1} - a_{<i})}{e_i} (\gamma' - \gamma) + \frac{1}{p_i} [(a_{<i+1} - a_{<i}) - (n_i - a_{<i}) p_i]. \quad (15)$$

Taking expectations conditional on $\{a_{<j+1}\}$ and then over all $a_{<j+1}$ except $a_{<i}$, we see that m'_i is an almost unbiased estimate of n_i . Further, squaring (15) and omitting the region: $0 \leq a_{<k} \leq h_k n_{k-1}$, where $h_k < 1 - q_1 \dots q_{k-1}$ and $h_k n_{k-1}$ is integral, all k , when taking expectations over $\{a_{<j+1}\}$, we find that

$$E[(m'_i - n_i)^2] = \frac{(n_i - \alpha_{<i})^2 p_i^2}{e_i^2} \frac{\gamma^2}{\sum_k \alpha_{<k} p_k} + (n_i - \alpha_{<i}) q_i / p_i + O(1) \quad (16)$$

for the limit process: $n_i \rightarrow \infty$ such that n_i / n_j is constant, all i, j . Besides being almost unbiased m'_i is also most efficient of the class of estimates m_i^* which satisfy obvious conditions.

It is worth examining the gain in information that constant catchability and knowledge of the effort bring. The information on n_1 provided by m'_1 represents a total gain as there is none when the effort is not known. Otherwise,

$$F'_k = \frac{\text{var}(n'_k)}{\text{var}(m'_k)} = \frac{1 + (\alpha_{<k+1} - \alpha_{<k}) / \alpha_{<k} p_k}{1 + (\alpha_{<k+1} - \alpha_{<k}) / \sum_l \alpha_{<l} p_l},$$

on replacing $(n_k - \alpha_{<k}) p_k$ by $\alpha_{<k+1} - \alpha_{<k}$ and e_k^2 / γ^2 by p_k^2 / q_k . In order to evaluate F'_k numerically, some assumptions must be made about the p_k and about the rate of immigration. First of all, suppose that $p_k = p$ for all k . Secondly, suppose that $\alpha_{<k+1} - \alpha_{<k}$ is constant for all k ; that is, the expected number of new individuals caught at each sample is constant. Taken in combination with the first assumption, this implies that $n_k - n_{k-1} = n_1 p$ for all k ; that is, the number of immigrants between consecutive samples is constant and equal to the expected size of the first sample. The formula for F'_k can be shown to be very insensitive

to the particular assumptions made about the rate of immigration, and we have merely chosen those which make the formula as simple as possible. It now becomes

$$F'_k = \frac{1 + [(k-1)p]^{-1}}{1 + [s(s-1)p/2]^{-1}}$$

and is tabulated below for $s = 5$ and three values of p . For $p \leq 0.001$, the formula $F'_k = s(s-1)/2(k-1)$ provides a very good approximation to the previous one.

$k \backslash p$	2	3	4	5
0.001	9.91	4.96	3.31	2.49
0.01	9.18	4.64	3.12	2.36
0.1	5.50	3.00	2.17	1.75

When we come to estimate the number of immigrants between samples, knowledge of the effort is doubly advantageous and results in a very considerable gain in information. For, $\text{var}(m'_k - m'_{k-1}) = \text{var}(m'_k) + \text{var}(m'_{k-1}) - 2\text{cov}(m'_k, m'_{k-1})$ and not only are the two variances smaller than their counterparts $\text{var}(n'_k)$ and $\text{var}(n'_{k-1})$, but their covariance is large and positive whereas $\text{cov}(n'_k, n'_{k-1})$ is negligible. In fact,

$$\text{cov}(m'_i, m'_j) = \frac{(\alpha_{<i+1} - \alpha_{<i})(\alpha_{<j+1} - \alpha_{<j})\gamma^2}{e_i e_j \sum \alpha_{<k} p_k} + O(1) \quad (i \neq j), \quad (17)$$

and, making the same simplifications as we did for F'_k , we find that

$$G'_h = \frac{\text{var}(n'_h - n'_{h-1})}{\text{var}(m'_h - m'_{h-1})} = 1 + \frac{1}{2p} \left[\frac{1}{h-1} + \frac{1}{h-2} \right] \quad (h = 3, \dots, s).$$

G'_h is tabulated below for $s = 5$ and three values of p .

$h \backslash p$	3	4	5
0.001	751	418	293
0.01	76.0	42.7	30.2
0.1	8.50	5.17	3.92

2.5. We mentioned in (I) that the capture-recapture model used by Hammersley (1953), when dealing with a large accumulation of data on the ringing of Alpine Swifts, contains a flaw. It is appropriate here to show how this comes about. Hammersley effectively made two postulates: that the overall likelihood could be taken simply as the product of the individual likelihoods for the captured birds; and that for any

such bird, its likelihood need only record its behaviour onwards from the sample in which it was first captured. For $s = 3$, the likelihood arrived at in this way is

$$(p_1 q_2 q_3)^{u_1} (p_1 p_2 q_3)^{u_{12}} (p_1 q_2 p_3)^{u_{13}} (p_1 p_2 p_3)^{u_{123}} (p_2 q_3)^{u_2} (p_2 p_3)^{u_{23}} p_3^{u_3}, \quad (18)$$

which may also be written

$$p_1^{a_1} p_2^{a_2} q_2^{a_{<3}-a_2} p_3^{a_3} q_3^{r-a_3},$$

or, for general s ,

$$\prod_i p_i^{a_i} q_i^{a_{<i+1}-a_i}.$$

The maximum likelihood estimate of p_i is $a_i/a_{<i+1}$. Dividing this into a_i to obtain an estimate of n_i , we get $a_{<i+1}$. In other words, the population size at any time is estimated by the number of individuals captured up to that time, which is plainly unsatisfactory. The reason why this model is at fault is indicated by comparing (18) with (1') and with (7) for $s = 3$. The latter can be written

$$p[\{u_w\} | \{a_{<i+1}\}] = \frac{a_{<2}!}{u_1! u_{12}! u_{13}! u_{123}!} (q_2 q_3)^{u_1} (p_2 q_3)^{u_{12}} (q_2 p_3)^{u_{13}} (p_2 p_3)^{u_{123}} \frac{(a_{<3}-a_{<2})!}{u_2! u_{23}!} q_3^{u_2} p_3^{u_{23}}. \quad (19)$$

A remark worth making here is that, even if (18) is replaced by a true density such as (19) and n_i is estimated by

$$\tilde{n}_i = a_i / \tilde{p}_i,$$

the sampling variance of $\tilde{n}_i$ is not only attributable to that of $\tilde{p}_i$ but also to that of a_i (and also to their covariance). This remark is of course implicit in §§ 2.3 and 2.4 where, for other reasons, we took

$$\tilde{n}_i = (a_{<i+1} - a_{<i}) / \tilde{p}_i + a_{<i}.$$

3. DEATH BUT NO IMMIGRATION

3.1. In this section, suffices i, j take all values from 1 to s and suffices k, l all values from 1 to $s-1$. p_i is now the conditional probability that an individual is captured at the i th sample, given that it is then alive. Let $a_{>k}$ be defined as the number of individuals caught after the k th sample, with a corresponding definition for $a_{k.>k}$.

Let χ_k be the conditional probability of an individual not being caught after the k th sample given that it is alive at the time of the k th sample. Then, for $s = 3$,

$$\begin{aligned} \chi_1 &= 1 - \phi_1 + \phi_1 q_2 (1 - \phi_2) + \phi_1 q_2 \phi_2 q_3 \\ &= 1 - \phi_1 p_2 - \phi_1 \phi_2 q_2 p_3, \\ \chi_2 &= 1 - \phi_2 + \phi_2 q_3, \\ &= 1 - \phi_2 p_3. \end{aligned}$$

Let π_0 denote the probability that an individual is never caught and π_i the probability that it is caught for the last time at the i th sample. Then, for $s = 3$,

$$\pi_0 = q_1 \chi_1, \quad \pi_1 = p_1 \chi_1, \quad \pi_2 = \phi_1 p_2 \chi_2, \quad \pi_3 = \phi_1 \phi_2 p_3$$

and

$$\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1,$$

which has an obvious probability interpretation.

The probability density of the u_w is

$$p[\{u_w\}] = \frac{n!}{(n-r)! \prod_w u_w!} \pi_0^{n-r} \prod_w P_w^{u_w},$$

where, for $s = 3$,

$$\begin{aligned} P_1 &= \pi_1, & P_3 &= q_1 q_2 \pi_3, \\ P_{13} &= p_1 q_2 \pi_3, \\ P_2 &= q_1 \pi_2, & P_{23} &= q_1 p_2 \pi_3, \\ P_{12} &= p_1 \pi_2, & P_{123} &= p_1 p_2 \pi_3. \end{aligned}$$

Hence,

$$p[\{u_w\}] = \frac{n!}{(n-r)! \prod_w u_w!} \pi_0^{n-r} \pi_1^{r-a_{>1}} \pi_2^{a_{>1}-a_{>2}} \pi_3^{a_{>2}-a_{>3}} p_1^{a_{>1}-a_{>2}} p_2^{a_{>2}-a_{>3}} q_1^{a_{>1}-a_{>2}} q_2^{a_{>2}-a_{>3}}.$$

3.2. Define $a_{>0} = r$ and $a_{>s} = 0$. Then, for general values of s ,

$$p[\{u_w\}] = \frac{n!}{(n-r)! \prod_w u_w!} \pi_0^{n-r} \prod_i \pi_i^{a_{>i-1}-a_{>i}} \prod_k p_k^{a_{>k}-a_{>k+1}} q_k^{a_{>k}-a_{>k+1}}. \quad (20)$$

Besides n , the original parameters were $\{\phi_k\}$ and $\{p_i\}$, a total of $2s-1$. We have now changed to π_0 , $\{\pi_i\}$ and $\{p_k\}$ subject to the constraints

$$\pi_0 + \sum_i \pi_i = 1, \quad \text{and} \quad \pi_0 p_1 - \pi_1 q_1 = 0.$$

The effective number of parameters is therefore $2s-2$, a reduction of one. The reason for this is that, if $p[\{u_w\}]$ is written as a function of $\{\phi_k\}$ and $\{p_i\}$, ϕ_{s-1} and p_s only appear in the combination $\phi_{s-1} p_s$ and are therefore really only one parameter. In other words, they are non-identifiable.

Maximizing (20) with Lagrangian multipliers λ_1, λ_2 for the two constraints, we find that

$$\begin{aligned} \lambda_1 &= 0, \quad \lambda_2 = \tilde{n}, \\ \tilde{p}_k &= \frac{a_{k,>k}}{a_{>k}}, \\ \tilde{n} \tilde{\pi}_i &= a_{>i-1} - a_{>i}. \end{aligned} \quad (21)$$

Let us define $\Phi_1 = 1$ and otherwise

$$\Phi_i = \phi_1 \phi_2 \dots \phi_{i-1},$$

the probability of survival up to the time of the i th sample. Then

$$\Phi_k p_k = \pi_k + p_k (\pi_{k+1} + \dots + \pi_s),$$

with obvious probability interpretation. Therefore

$$n \Phi_k = n \pi_k / p_k + (n \pi_{k+1} + \dots + n \pi_s).$$

Now let $\tilde{n}_k$ denote the estimate of $n \Phi_k$, which is the expected population size at the time of the k th sample. Then substituting the above estimates of $n \pi_i$,

$$\begin{aligned} \tilde{n}_k &= (a_{>k-1} - a_{>k}) / \tilde{p}_k + a_{>k} \\ \tilde{n}_k &= \frac{a_k a_{>k}}{a_{k,>k}}. \end{aligned} \quad (22)$$

and hence, substituting for $\tilde{p}_k$,

3.3. At this stage, the argument becomes so nearly parallel to that of §§ 2.3 and 2.4 that we can omit almost all of it. The starting point is the comparison of (21) and (22) with (6) and (8). Next, $n-r$ and $\{a_{>i-1} - a_{>i}\}$ are distributed multinomially with parameters n , π_0 and $\{\pi_i\}$. Therefore,

$$p[\{a_{>i-1}\}] = \frac{n!}{(n-r)! \prod_i (a_{>i-1} - a_{>i})!} \pi_0^{n-r} \prod_i \pi_i^{a_{>i-1} - a_{>i}}.$$

Hence

$$p[\{u_w\} | \{a_{>i-1}\}] = \frac{1}{\prod_w u_w!} \prod_i (a_{>i-1} - a_{>i})! \prod_k p_k^{a_{k.>k}} q_k^{a_{k.>k} - a_{k.>k}},$$

which should be compared with (7).

Let

$$n_k'' = \frac{(a_k + 1)(a_{>k} + 1)}{a_{k.>k} + 1} - 1.$$

Then n_k'' is almost unbiased for $n\Phi_k$ as n_k' was for n_k . There is, however, an interesting additional term in the expressions for $\text{var}(n_k'')$ and $\text{cov}(n_k'', n_l'')$ which was not present in the corresponding formulae for n_k' , n_l' . We find that

$$\text{var}(n_k'') = \frac{(n\Phi_k - \alpha_{>k})n\Phi_k q_k}{\alpha_{>k} p_k} + n\Phi_k(1 - \Phi_k) + O(1)$$

as $n \rightarrow \infty$, all other parameters remaining constant. The second term is the variance of the actual population size at the time of the k th sample. In the same way,

$$\text{cov}(n_k'', n_l'') = n\Phi_l(1 - \Phi_k) + o(1) \quad (k < l),$$

and the first term is the covariance of the actual population sizes at the times of the k th and l th samples.

In (I) we remarked that, if all $\phi_k = 1$, the n_k'' contain between them all of the information on n but that their evaluation requires 'differentiated' tagging.

3.4. If $q_i = e^{-e_i/\gamma}$ and the e_i are known, it is still true that

Hence

$$\tilde{n}\tilde{\pi}_i = a_{>i-1} - a_{>i}.$$

Let

$$\tilde{n}\tilde{\Phi}_i = (a_{>i-1} - a_{>i})/\tilde{p}_i + a_{>i}.$$

$$\gamma'' = \frac{\sum_k (a_{>k} - \frac{1}{2}a_{k.>k})e_k + \frac{1}{2}\bar{e}}{\sum_k a_{k.>k} + 1},$$

the almost unbiased estimate of γ and let m_i'' be the corresponding estimate of $n\Phi_i$. The terms $n\Phi_i(1 - \Phi_i)$ and $n\Phi_j(1 - \Phi_i)$ appear in the formulae for $\text{var}(m_i'')$, $\text{cov}(m_i'', m_j'')$, $i < j$, in addition to those corresponding to (16) and (17). But, as far as $F_k'' = \text{var}(n_k'')/\text{var}(m_k'')$ is concerned, the additional terms may be omitted because, although they are of the same order, $O(n)$, as the others, they are negligible by comparison. Therefore,

$$F_k'' = \frac{1 + (\alpha_{>k-1} - \alpha_{>k})/\alpha_{>k} p_k}{1 + (\alpha_{>k-1} - \alpha_{>k})/\sum_l \alpha_{>l} p_l}.$$

The simplifying assumptions are now that $p_k = p$ and $\alpha_{>k-1} - \alpha_{>k}$ is constant for all k . Taken in conjunction, these two assumptions imply that $n\Phi_k(1 - \phi_k) = n\Phi_s p_s$, that is the expected number of deaths between consecutive samples is constant and equal to the expected size of the last sample. With these simplifications,

$$F_k'' = F'_{s+1-k} \quad (k = 1, 2, \dots, s-1).$$

Also,
$$G_h'' = \frac{\text{var}(n_{h-1}'' - n_h'')}{\text{var}(m_{h-1}'' - m_h'')} = G_{s+2-h}' \quad (h = 2, \dots, s-1).$$

The experimenter will probably be more interested in estimating the survival probability ϕ_{h-1} than $n\Phi_{h-1} - n\Phi_h$, the expected number of deaths. Both n_h''/n_{h-1}'' and m_h''/m_{h-1}'' are biased estimates of ϕ_{h-1} but the biases are $O(n^{-1})$ and can easily be evaluated, if desired, using the δ -technique. The δ -technique also shows that, neglecting terms $O(n^{-1})$,

$$\frac{\text{var}(n_h''/n_{h-1}'')}{\text{var}(m_h''/m_{h-1}'')} = \frac{\text{var}(n_h'' - \phi_{h-1} n_{h-1}'')}{\text{var}(m_h'' - \phi_{h-1} m_{h-1}'')}$$

and, provided ϕ_{h-1} is not too much less than one, the latter ratio is well represented by G_h'' .

3.5. Throughout (I) and so far in this paper, we have been exclusively occupied with estimation and have said nothing about testing hypotheses. In this subsection, the likelihood-ratio method will be adapted for the purpose of testing hypotheses about the values of $\{\phi_k\}$ and $\{p_i\}$.

Consider any density of the form

$$p[\{u_w\}] = \frac{n!}{(n-r)! \prod_w u_w!} Q^{n-r} \prod_w P_w^{u_w},$$

where Q and the P_w are functions of parameters $\{\theta_t\}$ say. Let the log-likelihood of n and $\{\theta_t\}$ be $L_1 = \log p[\{u_w\}]$. Next, consider the conditional density

$$p[\{u_w\} | r] = \frac{r!}{\prod_w u_w!} \frac{\prod_w P_w^{u_w}}{(1-Q)^r},$$

and let the corresponding log-likelihood of $\{\theta_t\}$ be $L_2 = \log p[\{u_w\} | r]$.

Let $\tilde{L}_1$ and $\tilde{L}_2$ denote the maximum values of L_1 and L_2 . We show first that, to a good approximation, $\tilde{L}_1$ differs from $\tilde{L}_2$ only by an additive constant. Equating $\partial L_1 / \partial \theta_t$ to zero,

$$\frac{n-r}{Q} \frac{\partial Q}{\partial \theta_t} + \sum_w \frac{u_w}{P_w} \frac{\partial P_w}{\partial \theta_t} = 0. \quad (23)$$

Equating $\Delta_n L_1$ to zero,
$$\log n - \log(n-r) + \log Q = 0. \quad (24)$$

(24) gives $(n-r)/Q = r/(1-Q)$ and, substituting in (23), we have

$$\frac{r}{1-Q} \frac{\partial Q}{\partial \theta_t} + \sum_w \frac{u_w}{P_w} \frac{\partial P_w}{\partial \theta_t} = 0.$$

But this is seen to be the equation $\partial L_2 / \partial \theta_t = 0$. Therefore, the maximum likelihood estimates $\tilde{Q}$ and $\tilde{P}_w$ of Q and P_w are the same for L_1 and L_2 ; and for L_1 , $\tilde{n} = r/(1-\tilde{Q})$. To find $\tilde{L}_1$ and $\tilde{L}_2$, an approximation for $n!$ is required, and we shall assume that $n! = Ke^{-n}n^n$. (This approximation is equivalent to our practice of maximizing L_1 with respect to n by equating $\Delta_n L_1$ to zero rather than $\partial L_1 / \partial n$. For, if the exact identity $\Delta_n \log n! = \log n$ is replaced by the approximate one $\partial \log n! / \partial n = \log n$ and the latter is integrated, it leads directly to $n! = Ke^{-n}n^n$. Using this approximation instead of Stirling's, $n! = \sqrt{(2\pi)} e^{-n}n^{n+\frac{1}{2}}$, results in

an increase of $\frac{1}{2}$ in the maximum likelihood estimate of n , a difference which can be ignored as it is $O(n^{-1})$.) Hence,

$$\begin{aligned}\tilde{L}_1 &= \tilde{n}[\log \tilde{n} - 1] - (\tilde{n} - r)[\log(\tilde{n} - r) - 1] - \log \prod_w u_w! + (\tilde{n} - r) \log \tilde{Q} + \sum_w u_w \log \tilde{P}_w \\ &= r[\log r - 1] - \log \prod_w u_w! - r \log(1 - \tilde{Q}) + \sum_w u_w \log \tilde{P}_w \\ &= \tilde{L}_2 - \log K,\end{aligned}$$

which proves our assertion.

Let H_0 denote a hypothesis which reduces the number of degrees of freedom of $\{\theta_i\}$ by ν and H the alternative hypothesis allowing $\{\theta_i\}$ full freedom. Then, as L_2 has the properties required for maximum-likelihood large sample theory, we can say, using self-explanatory notation, that

$$x_2 = -2[\tilde{L}_{2,H_0} - \tilde{L}_{2,H}]$$

is approximately distributed as χ^2_ν . More precisely (Bartlett, 1953),

$$P[a \leq x_2 < b \mid r] = \int_a^b f(\chi^2_\nu) d\chi^2_\nu + O(r^{-1}) \quad \text{as } r \rightarrow \infty.$$

Let

$$x_1 = -2[\tilde{L}_{1,H_0} - \tilde{L}_{1,H}].$$

Then, since $\tilde{L}_{1,H} = \tilde{L}_{2,H} - K$ and, similarly, $\tilde{L}_{1,H_0} = \tilde{L}_{2,H_0} - K$,

$$x_1 = x_2.$$

Therefore, if P_r is the probability of r individuals,

$$P[a \leq x_1 < b] = \sum_{r \leq hn} P_r P[a \leq x_2 < b \mid r] + \sum_{r > hn} P_r P[a \leq x_2 < b \mid r],$$

where $0 < h < 1 - Q$ such that hn is integral. Now

$$\sum_{r \leq hn} P_r P[a \leq x_2 < b \mid r] \leq \sum_{r \leq hn} P_r = O(n^{-\frac{1}{2}} c^n) = o(n^{-1}),$$

where

$$c = \left(\frac{1-Q}{h}\right)^h \left(\frac{Q}{1-h}\right)^{1-h} < 1.$$

Also,

$$\begin{aligned}\sum_{r > hn} P_r P[a \leq x_2 < b \mid r] &= \sum_{r > hn} P_r \left[\int_a^b f(\chi^2_\nu) d\chi^2_\nu + O(r^{-1}) \right] \\ &= [1 - o(n^{-1})] \int_a^b f(\chi^2_\nu) d\chi^2_\nu + O(n^{-1}).\end{aligned}$$

Hence,

$$P[a \leq x_1 < b] = \int_a^b f(\chi^2_\nu) d\chi^2_\nu + O(n^{-1})$$

which is in the same form as the usual likelihood-ratio test.

There are two applications of this test for which the algebra has already been performed. In each case H denotes the hypothesis allowing complete freedom to the parameters $\{\phi_k\}$ and $\{p_i\}$ and it is therefore easy to find $\tilde{L}_{1,H}$. As the first example, consider $H_0: \phi_k = 1$ all k . $\tilde{L}_{1,H_0}$ can be found from (I) and $\nu = (2s-2) - s = s-2$. (If the test dictates that H_0 be accepted it will of course not mean that we believe it to be exactly true since this is

strictly impossible for an animal population. It will mean, rather, that the $\tilde{\phi}_k$ cannot be regarded as reliable estimates of the ϕ_k .) As the second example, consider $H_0: q_i = e^{-e_i/\gamma}$ where the e_i are known. This will test whether or not catchability is constant. $\tilde{L}_{1,H_0}$ can be found from § 3.4 and, again, $\nu = (2s - 2) - s = s - 2$.

Finally, a remark concerning hypothesis testing when there is immigration but no death. The above adaptation of the likelihood-ratio method is not applicable to the density (3). There is, however, a way round this difficulty. We can set up a formal model for immigration but no death which is the exact inverse of the model for death but no immigration. Namely, let the population be of size n at the time of the last sample and let $1 - \psi_k$ be the conditional probability that any individual immigrated between the $(k - 1)$ th and k th samples given that it was in the population at the time of the k th, $k = 2, 3, \dots, s$. This model yields just the same estimates as (3) and any likelihood-ratio test can be written down automatically from the corresponding one for death but no immigration. (For this model $p[\{u_w\}]$ is most quickly arrived at directly as a multinomial probability, but it can also be developed as a chain of conditional probabilities $P[S_1]P[S_2|S_1]\dots P[S_s|S_1, \dots, S_{s-1}]$ just as easily as for any other model. The main snag about this description of immigration is that it cannot be combined with stochastic death. To include both immigration and death, densities (3) and (20) must be 'combined', and we do this in the next and final section.)

4. IMMIGRATION AND DEATH

4.1. For $s = 3$, the generating function $E[\prod_w t_w^{u_w}]$ of $p[\{u_w\}]$ is clearly (cf. (2))

$$\begin{aligned} & (q_1\chi_1 + p_1\chi_1t_1 + p_1\phi_1p_2\chi_2t_{12} + p_1\phi_1q_2\phi_2p_3t_{13} + p_1\phi_1p_2\phi_2p_3t_{123} \\ & + q_1\phi_1p_2\chi_2t_2 + q_1\phi_1p_2\phi_2p_3t_{23} + q_1\phi_1q_2\phi_2p_3t_3)^{n_1} \\ & \times (q_2\chi_2 + p_2\chi_2t_2 + p_2\phi_2p_3t_{23} + q_2\phi_2p_3t_3)^{n_2-n_1} (q_3 + p_3t_3)^{n_3-n_2}, \end{aligned} \quad (25)$$

where χ_1 and χ_2 have been defined in § 3.1. The coefficient of $\prod_w t_w^{u_w}$ in (25) turns out to be a double sum of probabilities which only contracts to a single probability if $n_1 = n_2$ or $\phi_1 = 1$ and if $n_2 = n_3$ or $\phi_2 = 1$.

In general, $p[\{u_w\}]$ is an $(s - 1)$ -dimensional sum of probabilities which cannot be written as a single expression unless, in every interval between successive samples, there is either no immigration or no death.

4.2. The unwieldy form of $p[\{u_w\}]$ clearly rules out estimation by maximizing the likelihood, but there is an obvious pointer to an alternative method. In all cases when there is no immigration or no death, the method of maximum likelihood is equivalent to equating some of the class sizes to their expected values. We recall from (I) that, for a closed population, these are $\{a_i\}$ and r . It is easily shown that, for immigration but no death, they are $\{a_i\}$, $a_{<3}, \dots, a_{<s}$ and r , and, for death but no immigration, they are $\{a_i\}$, r and $a_{>1}, \dots, a_{>s-2}$. Therefore, when there is immigration and death, it seems eminently reasonable to equate

$$\{a_i\}, \quad a_{<3}, \dots, a_{<s}, \quad r, \quad a_{>1}, \dots, a_{>s-2} \quad (26)$$

to their expected values. We shall indicate how to solve the resulting equations by supposing that there are four samples, this being the minimum number from which the general solution may be inferred.

For $s = 4$, the aforesaid equations can be arranged as

$$a_1 = n_1 p_1, \quad (27 a)$$

$$a_2 = [n_1 \phi_1 + n_2 - n_1] p_2, \quad (27 b)$$

$$a_3 = [n_1 \phi_1 \phi_2 + (n_2 - n_1) \phi_2 + n_3 - n_2] p_3, \quad (27 c)$$

$$a_4 = [n_1 \phi_1 \phi_2 \phi_3 + (n_2 - n_1) \phi_2 \phi_3 + (n_3 - n_2) \phi_3 + n_4 - n_3] p_4, \quad (27 d)$$

$$r - a_{>1} = n_1 p_1 \chi_1, \quad (28 a)$$

$$a_{>1} - a_{>2} = [n_1 \phi_1 + n_2 - n_1] p_2 \chi_2, \quad (28 b)$$

$$a_{>2} - a_{>3} = [n_1 \phi_1 \phi_2 + (n_2 - n_1) \phi_2 + n_3 - n_2] p_3 \chi_3, \quad (28 c)$$

$$a_{<2} = n_1 p_1, \quad (29 a)$$

$$a_{<3} - a_{<2} = [n_1 \phi_1 q_1 + n_2 - n_1] p_2, \quad (29 b)$$

$$a_{<4} - a_{<3} = [n_1 \phi_1 \phi_2 q_1 q_2 + (n_2 - n_1) \phi_2 q_2 + n_3 - n_2] p_3, \quad (29 c)$$

where (27 a) reappears, for convenience, as (29 a).

Combining (27 a, b, c) with (28 a, b, c), we have $\chi_k = (a_{>k-1} - a_{>k})/a_k$ or $1 - \chi_k = a_{k.>k}/a_k$, $k = 1, 2, 3$. There is a recurrence relationship between successive χ_k which may be written

$$\phi_k q_{k+1} (1 - \chi_{k+1}) + \phi_k p_{k+1} = 1 - \chi_k \quad (k = 1, 2).$$

Hence

$$\phi_1 q_2 \frac{a_{2.>2}}{a_2} + \phi_1 p_2 = \frac{a_{1.>1}}{a_1}, \quad (30 a)$$

$$\phi_2 q_3 \frac{a_{3.>3}}{a_3} + \phi_2 p_3 = \frac{a_{2.>2}}{a_2}. \quad (30 b)$$

If we multiply (29 a) by $\phi_1 p_2$, add it to (29 b) and refer to (27 b), we obtain

$$a_{<3} - a_{<2} + \phi_1 p_2 a_{<2} = a_2$$

or

$$\phi_1 p_2 a_{<2} = a_{<2.2}. \quad (31 a)$$

Similarly, multiplying (29 a) by $\phi_1 \phi_2 p_3$, (29 b) by $\phi_2 p_3$ and adding them to (29 c),

$$\phi_2 p_3 (a_{<3} - a_{<2}) + \phi_1 \phi_2 p_3 a_{<2} = a_{<3.3}. \quad (31 b)$$

Equations (30) and (31) enable us to solve for p_2 , p_3 and ϕ_1 , ϕ_2 . Hence, from (27 b, c), we can estimate $n_1 \phi_1 + n_2 - n_1$ and $n_1 \phi_1 \phi_2 + (n_2 - n_1) \phi_2 + n_3 - n_2$, the expected population sizes at the times of the second and third samples. Also $n_3 - n_2$, the number of immigrants between these two samples. We note that (27 d) has not proved to be of any use.

In general, it is possible to solve for $p_2, \dots, p_{s-1}$ and hence for the expected population sizes at the times of all but the first and last samples. Also for the survival rate in all but the last interval and the number of immigrants in all but the first and last intervals. The remaining parameters cannot be estimated and it is fairly obvious that this would hold true of any method of estimation.

Since the distribution of the variables (26) is simply a combination of independent multinomial distributions, it is theoretically straightforward to find formulae for the variances and covariances of all the estimates by using the δ -technique. However, it would be too tedious to compile these formulae for the general case and we shall not attempt it here.

4.3. If $q_i = e^{-e_i \gamma}$ and the e_i are known, we can again appeal to moment equations for the purpose of estimation. In (I) we remarked that, for a closed population, maximizing the likelihood is equivalent to equating $\sum_i a_i e_i / p_i$ and r to their expected values. When there is immigration but no death the corresponding variables are $\sum_i a_i e_i / p_i$, $a_{<2} (= a_1)$, $a_{<3}$, ..., $a_{<s}$ and r , and when there is death but no immigration they are $\sum_i a_i e_i / p_i$, r and $a_{>1}$, ..., $a_{>s-2}$, $a_{>s-1} (= a_s)$. With both immigration and death we therefore equate $\sum_i a_i e_i / p_i$, $a_{<2}$, ..., $a_{<s}$, r , $a_{>1}$, ..., $a_{>s-1}$ to their expected values, thus providing $2s$ equations. These enable us to estimate all of the $2s$ unknowns γ , $\{n_i\}$ and $\{\phi_k\}$; except for $s = 2$ when the $a_1 e_1 / p_1 + a_2 e_2 / p_2$ equation is implied by the $a_{<2}$ and $a_{>1}$ equations and there are only three equations for the four unknowns.

Without actually performing the straightforward but tedious task of finding the variances of the estimates, we can be confident that knowledge of the effort provides very considerable gains in information.

I am indebted to the referee for his most helpful comments.

REFERENCES

- BARTLETT, M. S. (1953). Approximate confidence intervals. II. More than one unknown parameter. *Biometrika*, **40**, 306-17.
- DARROCH, J. N. (1958). The multiple-recapture census. I. Estimation of a closed population. *Biometrika*, **45**, 343-59.
- HAMMERSLEY, J. M. (1953). Capture-recapture analysis. *Biometrika*, **40**, 265-78.
- LESLIE, P. H., CHITTY, DENNIS & CHITTY, HELEN (1953). The estimation of population parameters from data obtained by means of the capture-recapture method. III. An example of the practical applications of the method. *Biometrika*, **40**, 137-69.

ON AN EXTENSION OF THE CONNEXION BETWEEN POISSON AND χ^2 DISTRIBUTIONS

BY N. L. JOHNSON
University College London

1. If a random variable x has a (central) χ^2 distribution with ν degrees of freedom then the probability density function of x is

$$p(x) = \frac{x^{\frac{1}{2}\nu-1} e^{-\frac{1}{2}x}}{2^{\frac{1}{2}\nu} \Gamma(\frac{1}{2}\nu)} \quad (x > 0).$$

The probability that x is less than a fixed value X (> 0) is

$$\Pr\{x < X\} = \int_0^X p(x) dx = \int_0^X \frac{x^{\frac{1}{2}\nu-1} e^{-\frac{1}{2}x}}{2^{\frac{1}{2}\nu} \Gamma(\frac{1}{2}\nu)} dx.$$

If ν is even, then integrating by parts we finally obtain

$$\begin{aligned} \Pr\{x < X\} &= 1 - e^{-\frac{1}{2}X} \left\{ 1 + \frac{\frac{1}{2}X}{1!} + \frac{(\frac{1}{2}X)^2}{2!} + \dots + \frac{(\frac{1}{2}X)^{\frac{1}{2}\nu-1}}{(\frac{1}{2}\nu-1)!} \right\} \\ &= e^{-\frac{1}{2}X} \left[\frac{(\frac{1}{2}X)^{\frac{1}{2}\nu}}{(\frac{1}{2}\nu)!} + \frac{(\frac{1}{2}X)^{\frac{1}{2}\nu+1}}{(\frac{1}{2}\nu+1)!} + \dots \right] \\ &= \Pr\{\xi \geq \frac{1}{2}\nu\}, \end{aligned} \tag{1}$$

where ξ is a Poisson variable with expected value $\frac{1}{2}X$.* This relationship between the Poisson and χ^2 distributions has proved of use in several cases (e.g. table 7 of Pearson & Hartley (1954) and particularly Cox (1955)).

In this paper an interesting extension, connecting the non-central χ^2 distribution with the distribution of the difference between two independent Poisson variables, is discussed. A similar relationship is found between non-central F and the difference between a negative binomial variable and an independent Poisson variable.

2. Let

$$x' = \sum_{i=1}^{\nu} (u_i + a_i)^2, \tag{2}$$

where $u_1, u_2, \dots, u_{\nu}$ are ν independent unit Normal variables and $a_1, a_2, \dots, a_{\nu}$ are constants with $\sum_{i=1}^{\nu} a_i^2 = \lambda$. Then the probability density function of x' is

$$p(x') = \sum_{j=0}^{\infty} \frac{e^{-\frac{1}{2}\lambda} (\frac{1}{2}\lambda)^j}{j!} \frac{x'^{\frac{1}{2}\nu+j-1} e^{-\frac{1}{2}x'}}{2^{\frac{1}{2}\nu+j} \Gamma(\frac{1}{2}\nu+j)} \quad (x' > 0).$$

This distribution of x' is called the non-central χ^2 distribution with ν degrees of freedom,

* ξ is used here (and ξ' later) to represent a random variable. This is contrary to usual practice, but the connexion between ξ and its mean $\frac{1}{2}X$ is similar to that between l (introduced later) and its mean $\frac{1}{2}\lambda$.

and non-centrality parameter λ . If ν is even, then, integrating with respect to x' , and using repeatedly the procedure by which (1) was obtained, we find

$$\Pr\{x' < X\} = \sum_{j=0}^{\infty} \frac{e^{-\frac{1}{2}\lambda} (\frac{1}{2}\lambda)^j}{j!} e^{-\frac{1}{2}X} \left[\frac{(\frac{1}{2}X)^{\frac{1}{2}\nu+j}}{(\frac{1}{2}\nu+j)!} + \frac{(\frac{1}{2}X)^{\frac{1}{2}\nu+j+1}}{(\frac{1}{2}\nu+j+1)!} + \dots \right]. \quad (3)$$

This can be written

$$\Pr\{x' < X\} = \sum_{j=0}^{\infty} \Pr\{l = j\} \Pr\{\xi \geq \frac{1}{2}\nu + j\},$$

where ξ is, as in § 1, a Poisson variable with expected value $\frac{1}{2}X$, and l is a Poisson variable with expected value $\frac{1}{2}\lambda$. If we further suppose that ξ and l are independent, (3) is equivalent to

$$\Pr\{x' < X\} = \Pr\{\xi - l \geq \frac{1}{2}\nu\}. \quad (4)$$

The relation (4) is new, so far as the author's knowledge goes, though Patnaik (1949) points out that $\Pr\{x' < X\}$ can be expressed as 'a double Poisson sum'.

3. If either side of (4) could be approximated it would, of course, provide an approximation to the other side.

It can be shown in straightforward fashion that if $k \geq 0$

$$\begin{aligned} \Pr\{\xi - l = k\} &= (X/\lambda)^{\frac{1}{2}k} e^{-(X+\lambda)} \sum_{j=0}^{\infty} \frac{(\frac{1}{2}\sqrt{[X\lambda]})^{k+2j}}{j! (k+j)!} \\ &= (X/\lambda)^{\frac{1}{2}k} e^{-(X+\lambda)} I_k(\frac{1}{2}\sqrt{[X\lambda]}), \end{aligned} \quad (5)$$

where $I_k(t)$ is the Bessel function, with imaginary argument, of order k (cf. Irwin, 1937; Skellam, 1946). (For $k \leq 0$ the only 'alteration' in the formula is the substitution of $-k$ for k as the order of the Bessel function. However, in evaluating (4), negative values of k cannot arise.)

With adequate tables the sums of terms such as those on the right-hand side of (5) could clearly be evaluated. However, Patnaik (1949) has shown that a simple (central) χ^2 approximation to the non-central χ^2 distribution gives quite good results.

It appears so far, therefore, that it is doubtful whether (4) helps us approximate non-central χ^2 integrals: it seems more likely that the converse process of approximating distributions of differences of independent Poisson variables by non-central χ^2 approximations will be useful. However, if neither $\frac{1}{2}X$ nor $\frac{1}{2}\lambda$ is too small, a Normal approximation (expected value $\frac{1}{2}(X - \lambda)$, standard deviation $\sqrt{[\frac{1}{2}(X + \lambda)]}$) to the distribution of $\xi - l$ should be justifiable. Using the natural continuity correction this leads to

$$\Pr\{x' < X\} \doteq \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{(X-\lambda-\nu+1)/\sqrt{2(X+\lambda)}} e^{-\frac{1}{2}x^2} dx. \quad (6)$$

Formula (6) has been obtained by assuming ν is even, but considerations of continuity indicate that it will be also a good approximation when ν is odd.

It is to be expected that the accuracy of (6) will increase with λ (and also with X). Patnaik's first non-central χ^2 approximation, on the other hand, is more accurate for small values of λ , i.e. when departure from central χ^2 form is not great. Hence one may hope the two approximations will be complementary, Patnaik's approximation being used to approximate the distribution of the difference of two independent Poisson variables when λ is small, and (6) used conversely to approximate to the non-central χ^2 distribution when λ is larger. The relation (4) is thus used to provide approximations for each of the two distributions concerned.

4. The numerical comparisons in this section, while not extensive, should be sufficient to indicate the usefulness of approximations based on (6). In so far as extensive tables of $I_k(t)$ are or become available, especially for greater values of k , exact probabilities, for either of the two distributions, could be evaluated exactly with some facility. Such tables might have other applications, e.g. in considering differences between experimental and control (background) counts. Such tables may indeed already be in existence and used for such purposes.

Approximate upper and lower $100\alpha\%$ points for the non-central χ^2 distribution may be obtained by solving

$$\frac{|X - \lambda - \nu + 1|}{\sqrt{2(X + \lambda)}} = u_\alpha \quad (7)$$

for X , where

$$\frac{1}{\sqrt{(2\pi)}} \int_{u_\alpha}^{\infty} e^{-\frac{1}{2}x^2} dx = \alpha.$$

Table 1 shows (i) some exact values of upper and lower 5% points taken from Patnaik's table 2, (ii) the errors in Patnaik's χ^2 approximation, (iii) the errors involved in determining X from (7) above and the results ((iv) and (v)) of two other approximations referred to below. In our Table 2a direct probabilities calculated from (6) are compared with exact values from Patnaik's table 6.

Table 1. Upper and lower 5% points of the non-central χ^2 distribution

(ii) Exact values. Errors (approx. - true) in: (ii) Patnaik's χ^2 approximation; (iii) approximation (7); (iv) approximation (7)'; (v) Pearson and Merrington's approximation.)

ν	λ	Upper 5%					Lower 5%				
		(i)	(ii)	(iii)	(iv)	(v)	(i)	(ii)	(iii)	(iv)	(v)
2	1	8.64	-0.01	0.92	0.57	-0.04	0.17	0.03	*	*	-0.09
	4	14.64	0.08	0.55	0.32	-0.06	0.65	0.29	-0.43	-0.28	-0.12
	16	33.06	0.29	0.28	0.16	-0.03	6.32	0.57	-0.25	-0.14	-0.02
	25	45.31	0.35	0.23	0.13	-0.03	12.08	0.60	-0.21	-0.12	-0.01
4	1	11.71	0.01	0.88	0.50	-0.02	0.91	0.02	-0.07	-0.06	-0.03
	4	17.31	0.07	0.57	0.33	-0.04	1.77	0.18	-0.24	-0.17	-0.06
	16	35.43	0.26	0.30	0.17	-0.03	7.88	0.48	-0.20	-0.13	-0.02
	25	47.61	0.33	0.25	0.14	-0.02	13.73	0.53	-0.17	-0.11	-0.01
7	1	16.00	0.01	0.83	0.47	-0.01	2.49	0.02	0.10	0.02	0.00
	4	21.23	0.05	0.59	0.34	-0.02	3.66	0.12	-0.07	-0.06	-0.02
	16	38.97	0.19	0.33	0.18	-0.02	10.26	0.38	-0.15	-0.09	-0.01
	25	51.06	0.28	0.26	0.14	-0.02	16.23	0.45	-0.14	-0.09	-0.01

* Formulae (7), (7)' give negative values in this case.

The relative accuracy of approximation (6) and (7), and Patnaik's χ^2 approximation (in Table 1a), varies with ν , λ and the limit of integration, X , roughly, but not quite, as anticipated in § 3. For small λ , Patnaik's is much the better of the two approximations. As λ increases the position becomes reversed. But the cross-over point (of about equal accuracy)

is at a considerably lower value of λ for the lower 5 % point than for the upper 5 % point. This is rather unexpected, though the *proportionate* accuracy of (7) is greater at the upper 5 % point as would be expected. (The results for $\nu = 7$, lower 5 % points, are rather anomalous, but the differences vary systematically, giving no indication of incorrect values.)

Table 2a. *Probability that non-central χ^2 exceeds upper 5 % point of corresponding central χ^2 distribution*

((i) Exact values (from Patnaik); (ii) approximation (6).)

$\lambda \backslash \nu$	2		4		8		12		20	
	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)
2	0.234	0.226	0.416	0.412	0.719	0.715	0.885	0.878	0.983	0.981
3	0.195	0.193	0.357	0.352	0.665	0.651	0.841	0.837	0.974	0.971
4	0.171	0.175	0.320	0.316	0.605	0.601	0.803	0.799	0.963	0.961
6	0.146	0.152	0.270	0.266	0.531	0.526	0.738	0.735	0.940	0.938
8	0.131	0.136	0.238	0.235	0.477	0.472	0.685	0.681	0.916	0.914
12	0.113	0.115	0.198	0.197	0.402	0.395	0.601	0.596	0.866	0.866
20	0.096	0.101	0.158	0.159	0.315	0.310	0.489	0.483	0.775	0.775

Table 2b. *Comparison of approximations to the non-central χ^2 probability integral*

((i) Exact values (from Patnaik); (ii) Patnaik's Normal approximation; (iii) approximation (6); (iv) approximation (6) modified by (6)'; (v) Pearson and Merrington's approximation.)

X	ν	λ	(i)	(ii)	(iii)	(iv)	(v)
30	16	32	0.0609	0.0638	0.0634	0.0632	0.0628
60	16	32	0.8316	0.8320	0.8310	0.8308	0.8320
36	24	24	0.1567	0.1515	0.1577	0.1577	0.1563
72	24	24	0.9667	0.9686	0.9644	0.9665	0.9667

5. Patnaik also gives a Normal approximation, which can be obtained by applying the relationship

$$\sqrt{(2\chi^2)} - \sqrt{(2\nu - 1)} \doteq \text{unit Normal variable}$$

to his χ^2 approximation. This leads to

$$\Pr\{x' < X\} \doteq \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^L e^{-\frac{1}{2}x^2} dx$$

with

$$L = \sqrt{\frac{2X(\nu + \lambda)}{\nu + 2\lambda}} - \sqrt{\left(\frac{2(\nu + \lambda)^2}{\nu + 2\lambda} - 1\right)}. \quad (8)$$

A few numerical comparisons of (6) and (8), using values in Patnaik's table 3, are shown in Table 2b. The two approximations appear to be of about equal accuracy, so far as is indicated by these four comparisons.

The limit of the integral in (6) is of the form

$$\frac{X - A}{\sqrt{(BX + C)}}, \quad (9)$$

while in (8) the limit is of form

$$D\sqrt{X - E}, \quad (10)$$

where A, B, C, D, E are constants depending on ν and λ . (9) is equal to zero when

$$X = A = \lambda + \nu - 1,$$

(10) equals zero when

$$X = (E/D)^2 = \lambda + \nu - \frac{1}{2}(\nu + 2\lambda)/(\nu + \lambda).$$

The latter value lies between $\lambda + \nu - \frac{1}{2}$ and $\lambda + \nu - 1$; hence the differences between the median non-central χ^2 given by the two approximations must be less than $\frac{1}{2}$, and approximation (7) will always give a smaller median value than Patnaik's Normal approximation.

The forms of expressions (9) and (10) indicate that good agreement between the two approximations cannot be expected over the whole range of values of X . When X is large, for example, the ratio of (10) to (9) is approximately

$$D\sqrt{B} = \sqrt{\left(\frac{2(\nu + \lambda)}{\nu + 2\lambda} \times 2\right)} = 2\sqrt{\frac{\nu + \lambda}{\nu + 2\lambda}} > \sqrt{2}.$$

Hence for upper percentage points we would expect (6) (or equivalently (7)) to give values of X considerably larger than those given by Patnaik's Normal approximation. Conversely, the latter should give larger values for $\Pr\{x' < X\}$. Table 2*b* indicates that this is the case, although the differences in values of $\Pr\{x' < X\}$ are not remarkably large. This may be because the values of X involved are not particularly large.

6. Equation (4) is exact. The accuracy of approximation (6) is essentially the accuracy of a Normal approximation to the distribution of the difference between two independent Poisson variables. This accuracy should be rather greater than that of a Normal approximation to either Poisson distribution separately, especially if $X \div \lambda$. The $\sqrt{\beta_1}$ skewness coefficient of the distribution of $\xi - l$ is

$$\frac{\sqrt{2}(X - \lambda)}{(X + \lambda)^{\frac{3}{2}}}.$$

A natural improvement to the Normal approximation may be obtained by adding the next term in the Gram-Charlier expansion, viz.

$$\frac{1}{3\sqrt{2}} \frac{X - \lambda}{(X + \lambda)^{\frac{3}{2}}} \left[\frac{(X - \lambda - \nu + 1)^2}{2(X + \lambda)} - 1 \right] \frac{1}{\sqrt{(2\pi)}} \exp \left[-\frac{(X - \lambda - \nu + 1)^2}{4(X + \lambda)} \right] \quad (6')$$

to the right-hand side of (6). The results of adding this correction to the values in column (iii) of Table 2*b* are shown in column (iv) of this table.

Another method of seeking an improvement on (6) is to approximate to each Poisson variable by a multiple of a χ^2 variable such that the two variables have the same mean and standard deviation. ξ, l are then replaced by $\frac{1}{2}\chi_X^2, \frac{1}{2}\chi_\lambda^2$, where χ_X^2, χ_λ^2 denote central χ^2 variables with X, λ degrees of freedom respectively.

From (4), we then obtain, using the continuity correction

$$\Pr\{x' < X\} \div \Pr\{\chi_X^2 - \chi_\lambda^2 > \nu - 1\}. \quad (11)$$

In general this result would not be useful as a basis for computation, since the distribution of $\chi_X^2 - \chi_\lambda^2$ is not simple to handle. If, however, one (preferably both) of the quantities X and λ are even integers, then the probability density function of $\chi_X^2 - \chi_\lambda^2$ can be obtained as a simple finite sum. If $\min(X, \lambda) = 2$ there is only one term in the expansion, if $\min(X, \lambda) = 4$ there are two terms, and so on.

Just as approximation (6) was modified by (6)' to allow for the skewness of distribution of $\xi - l$, we can also modify equation (7) for the percentage points of the non-central χ^2 distribution. Using the initial terms in the inverse Cornish-Fisher (1937) expansion we obtain

$$X - \lambda = \frac{(\nu - 1)(X + \lambda) - u_\alpha \sqrt{2(X + \lambda)^{\frac{3}{2}}}}{X + \lambda + \frac{1}{6}(u_\alpha^2 - 1)}. \quad (7)'$$

This is a quartic equation in X , which could be solved by an iterative method. Computation is rather simpler, however, if (7)' is used to obtain the value of $X - \lambda$ and hence the values of X and λ , corresponding to a given value of $X + \lambda$. From the table of values of X and λ so obtained the value of X , corresponding to a given λ , can be found by interpolation. Column (iv) of Table 1 gives the errors in values of X obtained by this method.

Columns (v) of Tables 1 and 2b give errors in values of X , and computed probabilities, respectively, obtained from a new approximation due to E. S. Pearson and Maxine Merington, which is described in a Note at the end of this paper.

7. We now consider the non-central F distribution. This is defined as the distribution of the non-central F -variable

$$F' = \frac{(\text{non-central } \chi^2 \text{ with } \nu \text{ degrees of freedom})/\nu}{(\text{central } \chi^2 \text{ with } \nu_0 \text{ degrees of freedom})/\nu_0},$$

the numerator and denominator being independent. The denominator will be written, for brevity, as v/ν_0 . Using (4), the conditional probability that F' does not exceed X/ν , given the value of v , is

$$\begin{aligned} \Pr\{F' < X/\nu \mid v\} &= \Pr\{x' < Xv/\nu_0\} \\ &= \Pr\{\xi_v - l \geq \tfrac{1}{2}\nu\} \end{aligned} \quad (12)$$

(if ν is even), where ξ_v is a Poisson variable, independent of l , with expected value $\frac{1}{2}Xv/\nu_0$. x' and l have the same meaning as in earlier sections. Averaging (12) over the distribution of v , we find the unconditional probability

$$\Pr\{F' < X/\nu\} = \Pr\{\xi' - l > \tfrac{1}{2}\nu\}, \quad (13)$$

where ξ' is a random variable, independent of l , with distribution obtained by averaging the Poisson distribution of ξ_v over the distribution of v . Thus

$$\begin{aligned} \Pr\{\xi' = j\} &= \frac{1}{2^{\frac{1}{2}\nu_0}\Gamma(\frac{1}{2}\nu_0)} \int_0^\infty v^{\frac{1}{2}\nu_0-1} e^{-\frac{1}{2}v} \frac{(\frac{1}{2}Xv/\nu_0)^j}{j!} e^{-\frac{1}{2}Xv/\nu_0} dv \\ &= \frac{1}{2^{\frac{1}{2}\nu_0}\Gamma(\frac{1}{2}\nu_0)} \frac{(X/2\nu_0)^j}{j!} \int_0^\infty v^{\frac{1}{2}\nu_0+j-1} e^{-\frac{1}{2}v(1+X/\nu_0)} dv \\ &= \frac{\Gamma(\frac{1}{2}\nu_0+j)}{\Gamma(\frac{1}{2}\nu_0)j!} \left(\frac{X}{\nu_0}\right)^j \left(1 + \frac{X}{\nu_0}\right)^{-\frac{1}{2}\nu_0-j} \quad (j = 0, 1, 2, \dots), \end{aligned}$$

so that ξ' is a negative binomial variable with integral probabilities given by terms in the expansion of

$$[(1 + X/\nu_0) - X/\nu_0]^{-\frac{1}{2}\nu_0}.$$

Equation (12) implies therefore that the distribution of non-central F can be expressed in terms of the distribution of the difference between a negative binomial variable (ξ') and an independent Poisson variable (l).

The expected value and variance of ξ' are

$$\frac{1}{2}\nu_0 \frac{X}{\nu_0} = \frac{1}{2}X$$

and

$$\frac{1}{2}\nu_0 \frac{X}{\nu_0} \left(1 + \frac{X}{\nu_0}\right) = \frac{1}{2}X \left(1 + \frac{X}{\nu_0}\right)$$

respectively. Hence the analogue of approximation (6) is

$$\Pr\{F' < X/\nu\} \doteq \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{(X-\lambda-\nu+1)/\sqrt{2(X+\lambda+X^2/\nu_0)}} e^{-\frac{1}{2}x^2} dx. \quad (14)$$

This differs from (6) only in the term X^2/ν_0 . Table 3 gives approximate values calculated using (14) compared with exact values and values of an approximation, using the central F distribution, obtained by Patnaik, taken from Patnaik's table 7. Table 3 also gives some further approximations analogous to those obtained for non-central χ^2 by the modification (6)' of (6). The corrective term to be added in the present case is

$$\frac{1}{3\sqrt{2}} \frac{X(1+2X/\nu_0)(1+X/\nu_0)-\lambda}{(X+\lambda+X^2/\nu_0)^{\frac{3}{2}}} \left[\frac{(X-\lambda-\nu+1)^2}{2(X+\lambda+X^2/\nu_0)} - 1 \right] \frac{1}{\sqrt{(2\pi)}} \exp \left[-\frac{(X-\lambda-\nu+1)^2}{4(X+\lambda+X^2/\nu_0)} \right]. \quad (14)'$$

Approximation (14), for the integral of the non-central F distribution, is not generally as accurate as Patnaik's central F approximation. This is in contrast to the situation for non-central χ^2 , and may correspond to the lower accuracy of the Normal approximation to the negative binomial distribution of ξ' compared with the Poisson distribution of ξ . The fall in accuracy of (14) as X increases is to be expected since the skewness of the negative binomial is greater than that of the Poisson distribution when X is large.

The exact distribution of $\xi' - l$ can be expressed in terms of the confluent hypergeometric function

$$M(a, c, x) = \frac{\Gamma(c)}{\Gamma(a)} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)}{\Gamma(c+j)j!} x^j. \quad (15)$$

We have for $k \geq 0$

$$\begin{aligned} \Pr\{\xi' - l = k\} &= \sum_{j=0}^{\infty} e^{-\frac{1}{2}\lambda} \frac{(\frac{1}{2}\lambda)^j}{j!} \frac{\Gamma(\frac{1}{2}\nu_0 + k + j)}{\Gamma(\frac{1}{2}\nu_0) \Gamma(k + j + 1)} \left(\frac{X}{\nu_0}\right)^{k+j} \left(1 + \frac{X}{\nu_0}\right)^{-\frac{1}{2}\nu_0 - k - j} \\ &= \frac{\Gamma(\frac{1}{2}\nu_0 + k)}{\Gamma(\frac{1}{2}\nu_0) k!} \left(\frac{X}{\nu_0}\right)^k \left(1 + \frac{X}{\nu_0}\right)^{-\frac{1}{2}\nu_0 - k} e^{-\frac{1}{2}\lambda} M(\frac{1}{2}\nu_0 + k, k + 1, \frac{1}{2}X\lambda/(X + \nu_0)). \end{aligned} \quad (16)$$

Direct evaluation of (16) would be facilitated by appropriate tables of the confluent hypergeometric function. The tables of Rushton & Lang (1954) might be useful, though some recurrence relations would be needed to cover the larger values of k .

For the inverse problem of calculating significance limits for F' , we can use an equation similar to (7). It is interesting to note that the quadratic equation (in X) derived from (7) can be changed into the corresponding equation (in $X = \nu \times$ (significance limit of F')) simply by reducing the coefficient of X^2 by $2u_a^2/\nu_0$.

8. In this paper it has been shown that the relationship between central χ^2 and Poisson variables can be extended in two ways. The cumulative distribution function of a non-central χ^2 can be expressed in the form of the upper tail of the distribution of

(a) (Poisson variable)–(Poisson variable), provided the number (ν) of degrees of freedom is even.

Also the cumulative distribution function of a non-central F can be expressed as the upper tail of the distribution of

(b) (Negative binomial variable)–(Poisson variable), again provided the number (ν) of degrees of freedom of the non-central χ^2 in the numerator is even.

Table 3. *Values of the non-central F integral, $\Pr\{F' < X/\nu\}$*

((i) Exact values; (ii) Patnaik's central F approximation;
(iii) approximation (14); (iv) approximation (14) modified by (14)'.)

ν	ν_0	λ	X/ν	(i)	(ii)	(iii)	(iv)
3	10	4	3.708	0.745	0.752	0.755	0.736
		4	6.552	0.918	0.919	0.889	0.901
		16	3.708	0.206	0.203	0.219	0.213
		16	6.552	0.517	0.520	0.544	0.517
3	20	4	3.098	0.700	0.706	0.711	0.694
		4	4.938	0.887	0.889	0.873	0.879
		16	3.098	0.126	0.119	0.129	0.130
		16	4.938	0.347	0.350	0.364	0.350
5	10	6	3.326	0.731	0.731	0.746	0.726
		6	5.636	0.914	0.913	0.886	0.903
		24	3.326	0.158	0.157	0.165	0.165
		24	5.636	0.461	0.463	0.504	0.461
5	20	6	2.711	0.664	0.665	0.680	0.660
		6	4.103	0.870	0.869	0.859	0.862
		24	2.711	0.069	0.064	0.068	0.072
		24	4.103	0.245	0.244	0.257	0.249
8	10	9	3.072	0.714	0.715	0.743	0.718
		9	5.057	0.908	0.909	0.882	0.893
		36	3.072	0.119	0.117	0.118	0.125
		36	5.057	0.408	0.409	0.451	0.409
8	30	9	2.226	0.578	0.581	0.596	0.576
		9	3.173	0.813	0.815	0.813	0.809
		36	2.226	0.017	0.014	0.015	0.017
		36	3.173	0.088	0.085	0.086	0.090

For completeness it is interesting to inquire whether the upper tails of the distributions of

(c) (Poisson variable)–(Negative binomial variable),

(d) (Negative binomial variable)–(Negative binomial variable) correspond to the cumulative distribution functions of any continuous variables of well-known types.

The second variable in (a) or (b) can be changed into a negative binomial variable by replacing the fixed expected value, $\frac{1}{2}\lambda$, by a random variable with the distribution of a multiple of (central) χ^2 . In the special case when this χ^2 has the same number (ν) of degrees of freedom as the original non-central χ^2 this would correspond to a component of variance analysis of variance model (e.g. Johnson, 1948) as opposed to the 'parametric' model represented by a fixed λ . In this situation the non-central χ^2 becomes a central χ^2 with the same number (ν) of degrees of freedom.

Hence, if ν is even, (c) is then associated with a central χ^2 distribution with ν degrees of freedom. This is, however, only a special case of (c) and, even in this case, the original formula (1) provides a more useful result. When $\frac{1}{2}\lambda$ is a multiple of a central χ^2 with arbitrary degrees of freedom (ν') there seems to be no simple continuous distribution associated with (c). But if $\nu' < \nu$ we might regard our modified non-central χ^2 , x' , as arising from (2) in a mixed type of model, for which the a_i 's are random variables but are not linearly independent. Then (c) would provide (for even ν) the discontinuous counterpart.

Similarly, (d) can be associated, in a special case, with a central F -distribution with an even number (ν) of degrees of freedom in the numerator. Again, this is only a special case of (d), and the relation

$$\Pr\{F' < X/\nu\} = \Pr\{\xi' \geq \frac{1}{2}\nu\}$$

obtained by putting $\lambda = 0$ in (b), would be preferable for most purposes. The remarks about possible interpretations of the general form of (c) in terms of a mixed form of model apply to (d) also.

9. Finally we repeat the procedure used in obtaining (12), but replacing the central $\chi_{\nu_0}^2$ by a non-central χ^2 , with ν_0 degrees of freedom and non-centrality parameter λ_0 , which we will denote by $\chi_{\nu_0}^{\prime 2}$.

Since the probability density function of $\chi_{\nu_0}^{\prime 2}$ can be written

$$p(\chi_{\nu_0}^{\prime 2}) = \sum_{j=0}^{\infty} e^{-\frac{1}{2}\lambda_0} \frac{(\frac{1}{2}\lambda_0)^j}{j!} p(\chi_{\nu_0+2j}^2) \chi_{\nu_0+2j}^2 = \chi_{\nu_0}^{\prime 2},$$

it follows that the averaged ξ_{ν} is now a mixture of negative binomial variables, having the probability generating functions

$$[(1 + X/\nu_0) - tX/\nu_0]^{-\frac{1}{2}(\nu_0+2j)} \quad (17)$$

with probabilities $e^{-\frac{1}{2}\lambda_0} (\frac{1}{2}\lambda_0)^j / j!$ respectively ($j = 0, 1, 2, \dots$). Hence if $F'' = (x'/\nu) / (\chi_{\nu_0}^{\prime 2}/\nu_0)$ represent an F -ratio based on *two* non-central χ^2 's then, if ν is even,

$$\Pr\{F'' < X/\nu\} = \sum_{j=0}^{\infty} e^{-\frac{1}{2}\lambda_0} \frac{(\frac{1}{2}\lambda_0)^j}{j!} \Pr\{\xi'_j - l \geq \frac{1}{2}\nu\}, \quad (18)$$

where ξ'_j is a negative binomial variable independent of l , having probability generating function (17). We note that ξ'_j has the same distribution as the random variable ξ' which appears in (13), with ν_0 replaced by $\nu_0 + 2j$ and X replaced by $X(\nu_0 + 2j)/\nu_0$. Since

$$\mathcal{E}(\xi'_j) = \frac{1}{2}X(1 + 2j/\nu_0)$$

and

$$\text{var}(\xi'_j) = \frac{1}{2}X(1 + X/\nu_0)(1 + 2j/\nu_0),$$

$\Pr\{\xi'_j - l \geq \frac{1}{2}\nu\}$ tends to 1 as j increases without limit. Hence, for a sufficiently large value of j , the remaining terms in the series in the right-hand side of (18) can be evaluated as a Poisson sum. A possibly useful way of evaluating (18) approximately would be to regard the right-hand side as a formula for the expected value of $\Pr\{\xi'_j - l \geq \frac{1}{2}\nu\}$ when j has

a Poisson distribution with mean $\frac{1}{2}\lambda_0$. The first term in an evaluation of this expected value by the method of statistical differentials, for example, would be $\Pr\{\xi'' - l \geq \frac{1}{2}\nu\}$, where ξ'' is a Poisson random variable with mean $\frac{1}{2}\lambda_0$; usually, of course, further terms would be needed.

10. We had, in the last paragraph of § 9, a 'mixture' of negative binomial variables (see Robbins & Pitman (1949) for formal use of the term 'mixture'). The concept of a 'mixture' of variables has a number of other applications in the present context. We can regard the non-central χ^2 distribution as a mixture of central χ^2 distributions, the negative binomial distribution as a mixture (with continuous weighting) of Poisson distributions and the non-central G -ratio distribution as a mixture of central G distributions. Here, central G_{ν, ν_0} is defined as the ratio of a central χ^2_ν to an independent central $\chi^2_{\nu_0}$; for non-central G the numerator becomes a non-central χ^2 , χ'^2_ν , say. A reasonable type of approximation to investigate is that obtained by using a mixture containing only two of the component distributions. This is a natural first extension of single component approximations of the type used by Patnaik for non-central χ^2 and non-central F .

We will find it convenient to use the following lemma (Jones, 1933).

LEMMA. If $cz_1^r + (1-c)z_2^r = k_r$ ($r = 1, 2, 3$) then z_1, z_2 must be roots of the equation

$$(k_2 - k_1^2)z^2 - (k_3 - k_1k_2)z + k_1k_3 - k_2^2 = 0.$$

This result follows from the equations

$$k_2 - k_1^2 = c(1-c)(z_1 - z_2)^2,$$

$$k_3 - k_1k_2 = c(1-c)(z_1 + z_2)(z_1 - z_2)^2,$$

$$k_1k_3 - k_2^2 = c(1-c)z_1z_2(z_1 - z_2)^2.$$

(i) *Non-central χ^2*

If χ'^2_ν be a non-central χ^2 with ν degrees of freedom and non-centrality parameter λ , we have

$$\mu'_1(\chi'^2_\nu) = \nu + \lambda, \quad \mu'_2(\chi'^2_\nu) = 2(\nu + 2\lambda) + (\nu + \lambda)^2,$$

$$\mu'_3(\chi'^2_\nu) = 8(\nu + 3\lambda) + 6(\nu + 2\lambda)(\nu + \lambda) + (\nu + \lambda)^3.$$

If y be a mixture of two central χ^2 's with ν_1, ν_2 degrees of freedom respectively, in the ratio $c : (1-c)$, then, equating the first three moments of y and χ'^2_ν we find

$$c\nu_1 + (1-c)\nu_2 = \nu + \lambda,$$

$$c\nu_1^2 + (1-c)\nu_2^2 = (\nu + \lambda)^2 + 2\lambda,$$

$$c\nu_1^3 + (1-c)\nu_2^3 = (\nu + \lambda)^3 + 6\lambda(\nu + \lambda) + 4\lambda.$$

Using the Lemma, ν_1, ν_2 are roots of

$$z^2 - 2(\nu + \lambda + 1)z + (\nu + \lambda)^2 + 2\nu = 0,$$

so

$$\nu_1 = \nu + \lambda + 1 + \sqrt{(2\lambda + 1)},$$

$$\nu_2 = \nu + \lambda + 1 - \sqrt{(2\lambda + 1)},$$

and

$$c = \frac{1}{2}(1 - (2\lambda + 1)^{-\frac{1}{2}}).$$

We can approximate similarly to the distribution of $a_1\chi_{w_1}^2 + a_2\chi_{w_2}^2$. We obtain equations

$$cv_1 + (1-c)v_2 = W,$$

$$cv_1^2 + (1-c)v_2^2 = W^2 + 2Y,$$

$$cv_1^3 + (1-c)v_2^3 = W^3 + 6WY + 4Z,$$

where

$$W = a_1w_1 + a_2w_2, \quad Y = a_1(a_1-1)w_1 + a_2(a_2-1)w_2,$$

$$Z = a_1(a_1-1)(2a_1-1)w_1 + a_2(a_2-1)(2a_2-1)w_2.$$

We find
$$\nu_1 = (Z + WY + \sqrt{[Z^2 + Y^3]})/Y, \quad \nu_2 = (Z + WY - \sqrt{[Z^2 + Y^3]})/Y,$$

$$c = \frac{1}{2}[1 - Z(Z^2 + Y^3)^{-\frac{1}{2}}].$$

it will usually be convenient to arrange that the variable is 'standardized' by making $a_1 + a_2 = 1$.

This result can be used in approximating to the distribution of modified t ,

$$t' = \frac{u \sqrt{(a_1w_1 + a_2w_2)}}{\sqrt{(a_1\chi_{w_1}^2 + a_2\chi_{w_2}^2)}},$$

where u is a unit Normal variable independent of the χ^2 's. Modified t is encountered in connexion with some criteria for testing differences between means of Normal populations with unknown variances (Welch (1938)). Applying the approximation of this section we have

$$\Pr\{t' < t_0\} \doteq c \Pr\{t_{\nu_1} < t_0 \sqrt{[\nu_1/(a_1w_1 + a_2w_2)]}\} + (1-c) \Pr\{t_{\nu_2} < t_0 \sqrt{[\nu_2/(a_1w_1 + a_2w_2)]}\}. \quad (19)$$

Non-central t' (u replaced by $u + \delta$) can similarly be expressed in terms of non-central t integrals.

(ii) Negative binomial

Let y be a negative binomial variable with probability generating function

$$(q - tp)^{-n}.$$

Then the mixture, in the ratio $c : (1-c)$, of two Poisson distributions with means m_1, m_2 will have the same first three moments as y if

$$cm_1 + (1-c)m_2 = np,$$

$$cm_1^2 + (1-c)m_2^2 = n(n+1)p^2,$$

$$cm_1^3 + (1-c)m_2^3 = n(n+1)(n+2)p^3.$$

m_1 and m_2 are the roots of

$$z^2 - 2(n+1)pz + n(n+1)p^2 = 0,$$

so

$$m_1 = p\sqrt{[n+1]}(\sqrt{[n+1]} + 1), \quad m_2 = p\sqrt{[n+1]}(\sqrt{[n+1]} - 1),$$

and

$$c = \frac{1}{2}(1 - (n+1)^{-\frac{1}{2}}).$$

It is interesting to note that with these values of m_1, m_2 and c , we have

$$cm_1^4 + (1-c)m_2^4 = n(n+1)^2(n+4)p^4$$

as compared with the value $n(n+1)(n+2)(n+3)p^4$ which would give exact agreement also with the fourth moment of y .

(iii) *Non-central F*

Two alternative approximations to the non-central F probability integral will be put forward here.

The first is based on the direct approximation to the non-central G distribution using (i) of this section to approximate the non-central χ^2 in the numerator. This leads immediately to the approximation to the distribution of non-central $G'_{\nu, \nu_0} = \chi'^2_{\nu} / \chi^2_{\nu_0}$ by the mixture in the ratio

$$(1 - (2\lambda + 1)^{-\frac{1}{2}}) : (1 + (2\lambda + 1)^{-\frac{1}{2}})$$

of the central G distributions of

$$G_{\nu_1, \nu_0}, \quad G_{\nu_2, \nu_0}, \quad \text{where} \quad \nu_1 = \nu + \lambda + 1 + \sqrt{(2\lambda + 1)}, \quad \nu_2 = \nu + \lambda + 1 - \sqrt{(2\lambda + 1)}.$$

Equivalently, $F'_{\nu, \nu_0} = (\chi'^2_{\nu} / \nu) / (\chi^2_{\nu_0} / \nu_0)$ is approximated by a mixture of $\nu_1 F_{\nu_1, \nu_0} / \nu$ and $\nu_2 F_{\nu_2, \nu_0} / \nu$ so that

$$\Pr\{F'_{\nu, \nu_0} < X/\nu\} \doteq \frac{1}{2}[1 - (2\lambda + 1)^{-\frac{1}{2}}] \Pr\{F_{\nu_1, \nu_0} < X/\nu_1\} + \frac{1}{2}[1 + (2\lambda + 1)^{-\frac{1}{2}}] \Pr\{F_{\nu_2, \nu_0} < X/\nu_2\}. \quad (20)$$

The second approximation is based on (13). Using (ii) of this section we have

$$\Pr\{\xi' - l \geq \frac{1}{2}\nu\} \doteq \frac{1}{2}[1 - (\frac{1}{2}\nu_0 + 1)^{-\frac{1}{2}}] \Pr\{\xi_a - l \geq \frac{1}{2}\nu\} + \frac{1}{2}[1 + (\frac{1}{2}\nu_0 + 1)^{-\frac{1}{2}}] \Pr\{\xi_b - l \geq \frac{1}{2}\nu\}, \quad (21)$$

where ξ_a, ξ_b are Poisson variables, independent of l , with expected values

$$\frac{1}{2}X_a = X \sqrt{(\frac{1}{2}\nu_0 + 1)} (\sqrt{[\frac{1}{2}\nu_0 + 1] + 1}) / \nu_0, \quad \frac{1}{2}X_b = X \sqrt{(\frac{1}{2}\nu_0 + 1)} (\sqrt{[\frac{1}{2}\nu_0 + 1] - 1}) / \nu_0$$

respectively. Proceeding further we use (4) to express (21) in terms of non-central χ^2 integrals. We then use (ii) of this section to obtain an approximation to the non-central F integral in terms of four central χ^2 integrals. This is

$$\begin{aligned} \Pr\{F'_{\nu, \nu_0} < X/\nu\} \doteq & \frac{1}{2}[1 - (\frac{1}{2}\nu_0 + 1)^{-\frac{1}{2}}] [\frac{1}{2}(1 - (2\lambda + 1)^{-\frac{1}{2}}) \Pr\{\chi^2_{\nu_1} < X_a\} \\ & + \frac{1}{2}(1 + (2\lambda + 1)^{-\frac{1}{2}}) \Pr\{\chi^2_{\nu_2} < X_a\}] \\ & + \frac{1}{2}[1 + (\frac{1}{2}\nu_0 + 1)^{-\frac{1}{2}}] [\frac{1}{2}(1 - (2\lambda + 1)^{-\frac{1}{2}}) \Pr\{\chi^2_{\nu_1} < X_b\} \\ & + \frac{1}{2}(1 + (2\lambda + 1)^{-\frac{1}{2}}) \Pr\{\chi^2_{\nu_2} < X_b\}], \end{aligned}$$

where X_a, X_b have the values defined above, and

$$\nu_1 = \nu + \lambda + 1 + \sqrt{(2\lambda + 1)}, \quad \nu_2 = \nu + \lambda + 1 - \sqrt{(2\lambda + 1)}.$$

REFERENCES

- CORNISH, E. A. & FISHER, R. A. (1937). Moments and cumulants in the specification of distributions. *Rev. Inst. Int. Statist.* **5**, 307-22.
- COX, D. R. (1955). Some statistical methods connected with series of events. *J. R. Statist. Soc. B*, **17**, 129-64.
- IRWIN, J. O. (1937). The frequency distribution of the difference between two independent variables following the same Poisson distribution. *J. R. Statist. Soc.* **100**, 415-16.
- JOHNSON, N. L. (1948). Alternative systems in the analysis of variance. *Biometrika*, **35**, 80-7.
- JONES, H. G. (1933). A note on the n -ages method. *J. Inst. Actu.* **64**, 318-24.
- PATNAIK, P. B. (1949). The non-central χ^2 and F distributions and their applications. *Biometrika*, **36**, 202-32.
- PEARSON, E. S. & HARTLEY, H. O. (1954). *Biometrika Tables for Statisticians*, **1**. Cambridge University Press.
- ROBBINS, H. E. & PITMAN, E. J. G. (1949). Application of the method of mixtures to quadratic forms in Normal variates. *Ann. Math. Statist.* **20**, 552-60.
- RUSHTON, S. & LANG, E. D. (1954). Tables of the confluent hypergeometric function. *Sankhyā*, **13**, 377-411.
- SKELLAM, J. G. (1946). The frequency distribution of the difference between two Poisson variates belonging to different populations. *J. R. Statist. Soc.* **109**, 296.
- WELCH, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, **29**, 350-62.

Note on an approximation to the distribution of non-central χ^2

By E. S. PEARSON

Dr Johnson's comparison on pp. 352-63 above of approximations to the probability integral of the non-central χ^2 distribution suggests that it might be appropriate to put on record here another approximation which Mrs Maxine Merrington and I tried out a little while ago. If, following Johnson, the non-central χ^2 variable is denoted by x' , then we have

$$\left. \begin{aligned} \mathcal{E}(x') &= \nu + \lambda, & \sigma(x') &= \sqrt{\{2(\nu + 2\lambda)\}}, \\ \beta_1(x') &= \frac{8(\nu + 3\lambda)^2}{(\nu + 2\lambda)^3}, & \beta_2(x') &= 3 + \frac{12(\nu + 4\lambda)}{(\nu + 2\lambda)^2}. \end{aligned} \right\} \quad (1)$$

It follows that the β_1, β_2 point of the distribution must fall within a fairly narrow sector of the β_1, β_2 field lying between the straight lines

$$\beta_2 - \frac{2}{3}\beta_1 - 3 = 0 \quad (\text{when } \lambda/\nu \rightarrow 0),$$

$$\beta_2 - \frac{4}{3}\beta_1 - 3 = 0 \quad (\text{when } \nu/\lambda \rightarrow 0),$$

the former being the χ^2 or Type III line. This suggests that, in shape, the distribution of x' will never differ greatly from that of a χ^2 distribution having appropriately chosen parameters. Patnaik gave the approximating distribution the correct start at $x' = 0$ and the correct mean and standard deviation. It occurred to us that except in cases where the curve rises abruptly at the lower end, i.e. when both ν and λ are small, it might be better to approximate by using a χ^2 distribution with the correct mean, standard deviation and β_1 coefficient, but not necessarily exactly the right start.

This latter result is achieved by taking

$$x' \sim \frac{\chi_{\nu'}^2 - \nu'}{\sqrt{(2\nu')}} \times \sigma(x') + \mathcal{E}(x') = \frac{\nu + 3\lambda}{\nu + 2\lambda} \chi_{\nu'}^2 - \frac{\lambda^2}{\nu + 3\lambda}, \quad (2)$$

where $\chi_{\nu'}^2$ is distributed as a central χ^2 having fractional degrees of freedom given by

$$\nu' = \frac{8}{\beta_1(x')} = \frac{(\nu + 2\lambda)^3}{(\nu + 3\lambda)^2}. \quad (3)$$

Using equations (2) and (3) and interpolating for ν' in the table of percentage points of χ^2 (*Biometrika Tables for Statisticians*, 1, table 8) we obtained upper and lower 5% points for non-central χ^2 with the errors as shown in columns (v) of Johnson's table 1. Again, for the X, ν and λ values given in his table 2b, we obtained the values of the probability integral shown in column (v) of that table; in this case, the values of the parameters selected, the three-moment χ^2 approximation is better than the other approximations. The computation involved is straightforward, apart from the necessity of interpolating for fractional degrees of freedom, ν' .

It should, however, be emphasized that a search for accurate approximations to non-central statistics was not the main purpose of Dr Johnson's investigation.

THE χ^2 TEST FOR SMALL EXPECTATIONS IN CONTINGENCY TABLES, WITH SPECIAL REFERENCE TO ACCIDENTS AND ABSENTEEISM

By C. A. G. NASS

Netherlands Institute of Preventive Medicine, Leiden

1. INTRODUCTION

In the study of absenteeism we are often presented with a large number of small samples of sickness absences or accidents, specified for two or more subperiods, each sample being contributed by one worker, observed during the total period. We desire to know whether the absences, say, of individual workers can be regarded as random samples of such a population, given by the marginal totals, regardless of whether or not the marginal totals agree with the lengths of the subperiods or some other assumption about the distribution of absences in time.

Thus we are presented with an $(m \times n)$ -design, with $(m-1)(n-1)$ degrees of freedom:

$$\left. \begin{array}{ccccc|c} x_{11} & \dots & x_{1j} & \dots & x_{1m} & z_1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{i1} & \dots & x_{ij} & \dots & x_{im} & z_i \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{n1} & \dots & x_{nj} & \dots & x_{nm} & z_n \\ \hline y_1 & \dots & y_j & \dots & y_m & N \end{array} \right\}. \quad (1)$$

The joint distribution with which are are concerned is one that could arise, among other ways, if from a finite population, subdivided into m classes with $y_1, \dots, y_m$ elements, samples of $z_1, \dots, z_{n-1}$ elements are randomly drawn without replacement, the remainder being considered as the n th sample with z_n elements. In our particular problem the rows of (1) represent workers, the columns subperiods and the letters the number of absences or accidents of one worker in one subperiod.

Our test of homogeneity or independence within the contingency table will be based on the statistic

$$G = N \sum_{i,j} \{x_{ij}^2 y_j^{-1} z_i^{-1}\} - N. \quad (2)$$

The standard test is to set $\chi^2 = G, \quad \nu = (m-1)(n-1).$ (3)

It is well known that if none of the expectations of x_{ij} are too small, then under the null hypothesis of independence, G will be distributed approximately in a chi-squared distribution with a parameter ν equal to the number of degrees of freedom. When, however, the expectations are small the distribution of G will differ from that of χ^2 , partly because the former becomes noticeably discontinuous and partly because the moments of G , beyond the first, may be rather different from those of χ^2 .

With absenteeism it often happens that a number of expectations of x_{ij} are smaller than unity, because a number of workers have only one absence in the total period of observation. Haldane (1937) met with the same difficulty in the field of genetics. In his

case the rows of (1) represent litters, the columns coat-colours and the letters numbers of mice. He suggested using the normal distribution as an approximation to the distribution of G , i.e.

$$u = (G - E(G))/\sqrt{\text{var } G}, \quad (4)$$

where u is assumed to be approximately normally distributed about zero with unit standard deviation. However, this approximation must be poor if the coefficient of variation of G is not small, which occurs in the field of absenteeism as it does in genetics (Grüneberg & Haldane, 1937).

Vessereau (1958), in a paper restricted to the one-way classification (a special case of design (1)), considered the approximation of the distribution of cG by a chi-squared distribution with parameter ν , where c and ν are determined such that the first two moments of the former distribution are equal to those of the latter, i.e.

$$\chi^2 = cG, \quad c = 2E(G)/\text{var } G, \quad \nu = cE(G). \quad (5)$$

He also considered the approximation of the distribution of $a + G$ by a chi-squared distribution, with suitably chosen a and ν . He found that it was markedly worse than in the case of (5), judged by the third and fourth moments.

It might be conceived intuitively that the approximation (5) must be the better one. If expectations are decreasing and the variance of G has ceased to be nearly twice its expectation, the general shape of the distribution of G might continue for a while to resemble that of some chi-squared distribution and it might continue to start at only a little above zero, because the values of $N(y_j z_i)^{-1}$ are not integral. Probably a still better approximation could be obtained by setting $\chi^2 = a + cG$, ensuring correct values for the first three moments. But the third moment of G for the general design (1) is at present unknown and even if determined, would probably be too unwieldy for practical use.

It is true that Vessereau considered the additional computations involved in determining the c and ν of (5) to be prohibitive. However, the greater labour involved is in calculating G , which must be performed in any case, rather than in calculating its variance, and we have suggested rules for simplifying the computation. Vessereau advised the use of the classical test (3), even in the presence of small expectations, suggesting that the 2% significance levels of the standard tables should be taken as 5% levels to allow for the greater variance, which was present in most of his purely numerical examples. Observing this advice, one could be fairly sure that the real probability of errors of the first kind would not exceed the nominal probability, but one would have to accept that the errors of the second kind would be grossly inflated. If it were true that small expectations tend to increase the variance of G , the inflation would only be moderate, but such a tendency does not generally exist. In most of the examples given in this paper, and taken from the practice of absenteeism, the variance of G is *less* than twice its expectation, and the same applies to the genetical examples of Grüneberg & Haldane. Besides, the errors of the second kind are not always on the 'safe' side. Which of the two kinds appears as the most dangerous depends on the particular problem considered. For these reasons we do not support Vessereau's advice.

In the present paper we shall consider methods of determining the significance of G , when expectations are small, based on Vessereau's approximation (5), although it was not favoured by its author. In so far as the resulting value of ν is beyond the range of existing

tables of the chi-squared distribution, we shall use R. A. Fisher's approximation, namely, assume that

$$u = \sqrt{(2\chi^2)} - \sqrt{(2\nu - 1)} = \sqrt{(2cG)} - \sqrt{(2\nu - 1)} \quad (6)$$

is normally distributed about zero with unit standard deviation. Comparisons between the true and approximate distribution of G , although limited to a few easily calculated cases, do suggest that the approximation is likely to be a good one in a very wide range of situations likely to be met in the analysis of accident and absenteeism data.

2. THE AVAILABLE TOOLS

The general design (1), and some of its special cases, constitute a handy set of tools for the study of accidents and absenteeism. For all of these the parameters c and ν are obtained in the same way (equation (5)), but the appropriate expressions for G , $E(G)$ and $\text{var } G$, to be given in this section, are different. For each of the tools we will give an appreciation of its special merits for certain jobs.

(i) *Case of $m \times n$ -fold table with $(m-1)(n-1)$ degrees of freedom*

This is the general design (1). Equation (2) gives the expression for G . That for $E(G)$ is

$$E(G) = \frac{(m-1)(n-1)N}{N-1}. \quad (7)$$

After publishing expressions for the moments of G for several special cases, Haldane (1939) found an expression for the variance of G for the general design. Dawson (1954) put Haldane's expression into the following shape, which is more suitable for computation:

$$\left. \begin{aligned} \rho &= \frac{(n-1)(N-n)}{N-1}, & \mu &= \frac{(m-1)(N-m)}{N-1}, \\ \sigma &= \frac{N\Sigma(z_i^{-1}) - n^2}{N-2}, & \tau &= \frac{N\Sigma(y_j^{-1}) - m^2}{N-2}, \\ \text{var } G &= \frac{2N}{N-3}(\rho - \sigma)(\mu - \tau) + \frac{N^2}{N-1}\sigma\tau. \end{aligned} \right\} \quad (8)$$

then

The merit of this tool is that it tests independence regardless of distribution in time and of distribution over workers. In this context 'independence' means that the set $(p_1, \dots, p_m)$ of probabilities of subperiods, for an arbitrary worker with z absences in the total period, is independent of z and of the worker. 'Distribution in time' means the distribution of absences over subperiods, regardless of the population of workers. Accordingly, this concept presupposes the existence of independence and it should be considered only after the data have successfully passed the test for independence, or if there are other strong reasons to take independence for granted. Under this pre-supposition the distribution in time is multinomial and assumptions about the values of $p_1, \dots, p_m$ can be tested with the classical χ^2 -test, since expectations will usually be large enough.

The distribution of absences over workers and its inverse, the distribution of workers over numbers of absences, lose practical importance if they are not constant over all periods with N absences in total. Analysis of such distributions should also be preceded by the test for independence, if the data are specified for subperiods.

(ii) Case of $m \times n$ -fold table with $m(n-1)$ or $(m-1)n$ degrees of freedom

These are limiting cases of the preceding one. In fact, if an $(m+1)$ th column is added to the design (1), with a column total Y tending to infinity and cell-frequencies X_i , such that the quotients X_i/Y tend to fixed numbers p_i , then the contribution of column (X_i) to G vanishes with probability one and (1) tends to a design with random row totals with, in the limit:

$$\left. \begin{aligned} G &= \sum_{i,j} x_{ij}^2 p_i^{-1} y_j^{-1} - N, \\ E(G) &= m(n-1), \quad \text{var } G = 2m(n-1) + (\sum p_i^{-1} + 2 - n^2 - 2n) \sum y_j^{-1}. \end{aligned} \right\} \quad (9)$$

The equations for the case with $(m-1)n$ degrees of freedom are symmetrical with (9). These results were found by Haldane (1937). In these cases independence is tested simultaneously with an assumed distribution of absences over either subperiods or workers. It is difficult to imagine a situation in the study of absenteeism where this model would be appropriate.

(iii) Case of one-way table

If we set $m = 1$ in case (ii) with $m(n-1)$ degrees of freedom, the design reduces to a discrete one-way distribution with n cells. Dropping the subscript j , the appropriate equations are:

$$\left. \begin{aligned} G &= \sum (x_i^2 / (N p_i)) - N, \quad E(G) = n - 1, \\ \text{var } G &= 2(n-1) - (n^2 + 2n - 2)/N + \sum (N p_i)^{-1}. \end{aligned} \right\} \quad (10)$$

These results were found by K. Pearson (1932).

Usually one or more parameters must be fitted to the sample distribution to find the p_i . To provide for this fact we suggest that $E(G)$ and $\text{var } G$, found with (10), be multiplied by $(n-k-1)/(n-1)$, if k parameters are fitted. This is the approximate effect of efficient parameter fitting when expectations are large. That it would apply also when expectations are small is only an unproved guess, but it seems better than no correction at all, which would certainly produce an inflation of the errors of the second kind.

It is often desirable to reserve the notation x_i , or x for the class values of the variate of the distribution tested. In that case the class frequencies might be denoted by n_i or n_x .

If the range of the distribution is infinite, it must be curtailed somewhere. Since departure from expectation often depends heavily on one or two very improbable sample values, a curtailment which excludes all expectations smaller than a few units is apt to produce a substantial loss of power. The merit of case (iii) is that it allows us to postpone the curtailment until expectations are well under one unit. Some results of § 4 below suggest that one could go as far as to include expectations of 0.05, if the number of cells is large and if there are also considerable numbers of cells with somewhat higher expectations.

(iv) Case of equal probabilities

If we set $p_i = n^{-1}$ in case (iii), equations (10) become:

$$G = n \sum (x_i^2) / N - N, \quad E(G) = n - 1, \quad \text{var } G = 2(n-1)(N-1)/N. \quad (11)$$

One merit of case (iv) is that it allows a correction for continuity, namely the subtraction of unity from the sample value of $\sum x_i^2$. It is seen from (11) that differences between adjacent values of G are proportionate to differences between corresponding values of $\sum x_i^2$. Thus the

proper correction should be to subtract from the sample value of Σx_i^2 half the distance to the preceding value. If one is subtracted from a cell-frequency, say x_1 , and added to a smaller cell-frequency, say x_2 , then Σx_i^2 is diminished by

$$x_1^2 + x_2^2 - (x_1 - 1)^2 - (x_2 + 1)^2 = 2(x_1 - x_2 - 1).$$

It follows that the distance to the preceding value is at least 2, and that it is exactly 2 if there is at least one pair of cell-frequencies that differ by two units. If the cell-expectations are small, it is most unlikely that none of the cells will be empty and that none will contain just two observations, unless all frequencies are either one or zero, in which case the value of Σx_i^2 is the smallest possible. In fact, it appears from the exact distributions of G , shown in § 4, with expectations 2, 1, 1, 0.5, and 0.2 respectively, that the possible values of Σx_i^2 proceed by steps of two units, up to a cumulative probability of at least 0.999. When expectations are not small, greater steps might occasionally be located in the critical part of the range, but then the number of possible values of G is large and the effect of discontinuity is small. It seems, therefore, that the subtraction of one unit from the sample value of Σx_i^2 is an effective correction where a correction is most needed.

It follows from equations (5) and (11) that, with correction for continuity:

$$\chi^2 = \frac{n(\Sigma x_i^2 - 1) - N^2}{N - 1}, \quad \nu = \frac{N(n - 1)}{N - 1}. \quad (12)$$

Thus another merit of the case (iv) test is that the computations involved are very simple. However, the most important merit is that it may be considered as an adaptation to small expectations of Fisher's test for Poisson distributions.

If we wish to test whether the observations $x_1, x_2, \dots, x_n$ are independent Poisson variates having a common expectation, the standard test, as first given by R. A. Fisher (1925), is to calculate

$$G = \frac{\Sigma(x_i - \bar{x})^2}{\bar{x}} = \frac{n\Sigma(x_i^2)}{N} - N.$$

This is the G of (11), the distribution of this criterion under the condition $\Sigma x_i = N$ being the same as the distribution of G considered above.

It is well known that Fisher's criterion discriminates powerfully between a Poisson distribution and a distribution with a variance noticeably greater than the mean, but poorly if the variance and the mean of the alternative distribution are also equal. Therefore, if the data have successfully passed the test of case (iv), it may not be superfluous to try the test of case (iii), fitting a Poisson distribution to the data and comparing the observed and expected cell-frequencies.

(v) *Case of n elements distributed over n cells with equal probabilities*

If we set $N = n$ in case (iv), equations (12) become

$$\chi^2 = \frac{n(\Sigma x_i^2 - n - 1)}{n - 1}, \quad \nu = n. \quad (13)$$

This gives a test particularly suited for testing an assumed even distribution of points on an interval, if we divide the interval into as many parts of equal length as there are points. It sometimes happens that the end-points of the interval of observation of a sequence of, say, accidents are not recorded. Then take the times of the first and the last accident as the

end-points and for the rest, disregard these two accidents. In fact, if $n+2$ points are randomly and evenly distributed on an interval (a, b) , then under the condition that the positions of the extreme points are a' and b' , the n other points are randomly and evenly distributed on the interval (a', b') .

3. METHODS OF COMPUTATION

The computations involved in case (i) of the preceding section are so extensive that it is a practical necessity to arrange them in some economical scheme. In the applications which we have in mind the number of rows (n) will usually be much greater than the number of columns (m) and most of the row totals (z_i) will have small integral values. Thus the quantity Σy_j^{-1} may be computed directly from the y_j , but for the computation of Σz_i^{-1} the z_i should then first be arranged in a frequency table. If n_z denotes the number of rows (workers) with $z_i = z$, then

$$\Sigma z_i^{-1} = \Sigma n_z z^{-1}.$$

It is economical to combine these computations with those for G . We shall give two schemes, one for $m > 2$ and one for $m = 2$.

Scheme for $m > 2$

- (1) Write the rows of design (1) on separate strips or on the upper margin of cards.
- (2) For each j , sort the strips according to x_{ij} . Note the frequency distribution of x_{ij} , its number, n , and its sum, y_j . Pool the frequency distributions for all j and note the sum, N and the sum of squares, Q .
- (3) Sort the strips according to the row total, z . Prepare a table with as many rows as there are different values of z and with $2m+5$ columns, under the headings,

$$(z) (z^{-1}) (n_z) \dots (s_j) \dots (s) \dots (q_j) \dots (q),$$

for each z writing under (n_z) the number of strips, under (s_j) the partial sum of the x_{ij} , under (s) the sum of the s_j , under (q_j) the partial sum of squares of the x_{ij} , under (q) the sum of the q_j .

- (4) Add the columns $(n_z), \dots, (s_j), \dots, (s)$ and (q) and check whether the sums are equal to $n, \dots, y_j, \dots, N$ and Q , already computed under step (2). Write under the (q_j) the product-sums

$$Q_j = (z^{-1}) \times (q_j).$$

Check whether

$$\Sigma Q_j = (z^{-1}) \times (q) \quad (\text{written under } (q)).$$

Write under (z^{-1}) the product-sum

$$\Sigma z_i^{-1} = (z^{-1}) \times (n_z).$$

- (5) Compute the sum Σy_j^{-1} directly from the y_j and the product-sum

$$\Sigma_{i,j} x_{ij}^2 y_j^{-1} z_i^{-1} = (y_j^{-1}) \times (Q_j).$$

- (6) Compute G , $E(G)$, $\text{var } G$, c , ν and χ^2 , using the equations (2), (7), (8) and (5).

It will be observed that the greater part of these computations can be done by mechanical operation, the squares of the x_{ij} being reproduced automatically, using a restricted number of master cards. The procedure is illustrated on a numerical example in § 5.

Scheme for $m = 2$

If there are only two columns, no separate strips are needed and the following simpler procedure may be followed:

(1) Prepare, by simple scoring, a contingency table of the rows of design (1), with the two cell-frequencies as entries. Denote the different values of the x_{i1} by x and those of the x_{i2} by $z - x$. The frequency distributions of x and $z - x$ are found by addition of rows and columns. Note the total number n and the sums y_1 and y_2 of x and $z - x$.

(2) Prepare a table with as many rows as there are different values of z and with five columns, headed

$$(z) (z^{-1}) (n_z) (x) (x^2).$$

For every z , write under (n_z) , (x) and (x^2) the total, the sum of x and the sum of x^2 over the diagonal of the contingency table with the corresponding value of z .

(3) Add the columns (n_z) and (x) and check whether the sums are equal to n and y_1 respectively. Check whether

$$N = (z) \times (n_z) = y_1 + y_2 \quad (\text{written under } (z)).$$

Write under (z^{-1}) and (x^2) the product-sums

$$\Sigma z_i^{-1} = (z^{-1}) \times (n_z) \quad \text{and} \quad Q = (z^{-1}) \times (x^2).$$

(4) For the computation of G use the equation

$$G = N(Q(y_1^{-1} + y_2^{-1}) - y_1 y_2^{-1}),$$

which is more convenient here than equation (2). Compute $E(G)$, $\text{var } G$, c , ν and χ^2 with equations (7), (8) and (5).

This simpler procedure is also illustrated in § 5.

4. SOME TEST CASES FOR THE ACCURACY OF THE APPROXIMATION

For designs restricted to two columns (case (i) of § 2 with $m = 2$), and row totals, z_i , not greater than two, the computation of exact distributions for G is not unduly heavy. Since the approximation presumably improves for $m > 2$ and larger row totals, such designs are particularly suited for test cases. Let n be the number of rows, n_1 of which have row total 1, the remaining n_2 having row total 2, then

$$n_1 + 2n_2 = y_1 + y_2 = N.$$

Of the cell-frequencies in column 1 with row total 2, let a be the number of those equal to 1 and b the number of those equal to 2. Then the numbers of the five possible types of row are:

Cell-frequencies	Number of rows
1, 1	a
2, 0	b
0, 2	$n_2 - a - b$
1, 0	$y_1 - a - 2b$
0, 1	$n_1 - y_1 + a + 2b$

The probability of a particular random sample is

$$y_1! y_2! 2^a / N!$$

and the number of possible samples with given a and b is

$$\frac{n_1!}{(y_1 - a - 2b)! (n_1 - y_1 + a + 2b)!} \frac{n_2!}{a! b! (n_2 - a - b)!}.$$

Thus the probability of a is

$$P(a) = \frac{y_1! y_2! n_1! n_2! 2^a}{N!} \sum_b \frac{1}{b! (n_2 - a - b)! (y_1 - a - 2b)! (n_1 - y_1 + a + 2b)!}. \tag{14}$$

The criterion G is found to depend on a only, since

$$\sum_{i,j} \frac{x_{ij}^2}{y_j z_i} = \frac{a}{2} \left(\frac{1}{y_1} + \frac{1}{y_2} \right) + \frac{4b}{2y_1} + \frac{4(n_2 - a - b)}{2y_2} + \frac{y_1 - a - 2b}{y_1} + \frac{n_1 - y_1 + a + 2b}{y_2} = 2 - \frac{aN}{2y_1 y_2}.$$

It follows by equation (2) that

$$G = N \left(1 - \frac{aN}{2y_1 y_2} \right).$$

Since a proceeds by steps of one, the value corrected for continuity is

$$G = N \left(1 - \frac{(2a + 1) N}{4y_1 y_2} \right). \tag{15}$$

Using equation (14), the exact distribution of a was computed for two designs of the type shown in Table 1.

Table 1

Design: $y_1 = y_2 = 20, n_1 = 10, n_2 = 15,$ $N = 40, n = 25$						Design: $y_1 = 280, y_2 = 20, n_1 = n_2 = 100,$ $N = 300, n = 200$					
a	$P(a)$	a	$P(a)$	a	$P(a)$	a	$P(a)$	a	$P(a)$	a	$P(a)$
0	0.00002	6	0.14071	12	0.01825	3	0.00002	9	0.05062	15	0.09893
1	0.00035	7	0.18906	13	0.00464	4	0.00011	10	0.09289	16	0.05102
2	0.00255	8	0.19840	14	0.00075	5	0.00060	11	0.14070	17	0.01975
3	0.01142	9	0.16285	15	0.00006	6	0.00252	12	0.17585	18	0.00541
4	0.03550	10	0.10378			7	0.00839	13	0.17997	19	0.00093
5	0.08108	11	0.05054			8	0.02276	14	0.14939	20	0.00007
Total 0.99996						Total 0.99993					

Application of equations (7), (8), (5), (15) and (6) gives the results shown in Table 2. The comparison of exact and approximate cumulative distributions of a , which by (15) are equivalent to those of G , is shown in Table 3.

Table 2

	Design: $y_1 = y_2 = 20$, $n_1 = 10$, $n_2 = 15$, $N = 40$, $n = 25$	Design: $y_1 = 280$, $y_2 = 20$, $n_1 = n_2 = 100$, $N = 300$, $n = 200$
$E(G)$ from (7)	24.615	199.67
var G from (8)	15.289	300.73
c from (5)	3.220	1.328
v from (5)	79.26	265.2
u from (15), (5) and (6)	$2.538\sqrt{(39-2a)} - 12.591$	$1.886\sqrt{(221-6a)} - 23.01$

Table 3

Design: $y_1 = y_2 = 20$, $n_1 = 10$, $n_2 = 15$						Design: $y_1 = 280$, $y_2 = 20$, $n_1 = n_2 = 100$					
a	Exact	Approx.	a	Exact	Approx.	a	Exact	Approx.	a	Exact	Approx.
0	0.000	0.001	8	0.659	0.624	4	0.000	0.000	12	0.494	0.495
1	0.000	0.002	9	0.822	0.834	5	0.001	0.001	13	0.674	0.676
2	0.003	0.008	10	0.926	0.937	6	0.003	0.004	14	0.824	0.825
3	0.014	0.024	11	0.976	0.983	7	0.012	0.013	15	0.923	0.923
4	0.050	0.062	12	0.995	0.997	8	0.034	0.036	16	0.974	0.973
5	0.131	0.141	13	0.999	1.000	9	0.085	0.086	17	0.994	0.993
6	0.272	0.275	14	1.000	—	10	0.178	0.181	18	0.999	0.998
7	0.461	0.460				11	0.319	0.319	19	1.000	1.000

In the first design the approximate 1% level of significance is correct, namely at $a = 2$. The exact 5% level is at $a = 4$, the exact and approximate probabilities of a being smaller than 5, being equal to 0.050 and 0.062 respectively. This design contains 20 cells with expectation 0.5. The second design contains 100 cells with an expectation of only 0.067. Yet the approximation is excellent for this design, the 1 and 5% levels of significance being correct. Our conclusion from these comparisons is that in the field of absenteeism there will be little occasion for serious errors, due to poor approximation, if case (i) of § 2 is applied. Vessereau (1958) has given exact distributions of G in three designs with N elements distributed over n equi-probable cells, namely for

$$N = 15, n = 5, \quad N = 16, n = 8, \quad N = 10, n = 10.$$

Such designs, apart from being the only cases of one-way classifications for which the computation of the exact distribution is practically possible, are of special importance, since the distribution is the same as that of Fisher's criterion, assuming that $\sum x_i = N$ is fixed (case (iv) of § 2). For our purpose, however, Vessereau's designs are not very critical, since their values of N and n , if presenting numbers of absences and of workers, seem far below anything that will be asked for in practice, whereas expectations smaller than $N/n = 1$ will frequently occur. One also wants to get some idea of the speed with which the approximation improves with increasing n in the important case where $n = N$, about its behaviour if n

varies with constant N and so forth. Unfortunately one has to take account of the amount of computation required; in particular, in finding the exact distributions. For designs with $n = N$ the work increases rapidly with N ; once it is done however for such a design, the additional work for designs with the same N and greater values of n is comparatively small. On such considerations we decided to take Vessereau's second and third designs and new designs with $N = 16$ and $n = 16, 32$ and 80 respectively. For these, we compare the exact distribution of G with its approximation according to (12).

Vessereau has described the following method of finding the exact distribution of G in a design of N elements distributed with equal probabilities over n cells: Let a *partition* of N over n be defined as an ordered set of n non-negative integers x_i , such that $\sum x_i = N$. Let $p_0, p_1, \dots, p_N$ be a sequence of non-negative integers. The union of all partitions with p_0 numbers equal to 0, ..., p_N numbers equal to N will be referred to as a *class*. All partitions of a class are equi-probable, since their probability is

$$\frac{n^{-N} N!}{(2!)^{p_2} \dots (N!)^{p_N}}.$$

The number of partitions of a class is $n!/(p_0! \dots p_n!)$ so that the probability of a class is

$$P(p_0 \dots p_N) = \frac{n^{-N} n! N!}{(2!)^{p_2} \dots (N!)^{p_N} p_0! \dots p_N!}. \quad (16)$$

Let Q be defined by

$$Q = \sum x_i^2 = p_1 + 4p_2 + \dots + N^2 p_N, \quad (17)$$

so that all partitions of a class have the same value of Q . Consider two designs with the same value of N , the first with $n = N$ and the second with a greater value of n . To every class of the first design there exists a class of the second design with the same values of $p_1, \dots, p_N$ and with a value of p_0 that is $n - N$ greater. Let q denote the value of p_0 under the first design, then the value of p_0 of the corresponding class under the second design is $q + n - N$ and corresponding classes have the same values of q and Q . Denote the probability of the union of all classes with given values of q and Q under the first design of $P_N(q, Q)$ and that of the corresponding union under the second design by $P_n(q, Q)$. It follows from (16) that

$$\left. \begin{aligned} P_N(q, Q) &= \frac{(N!)^2 N^{-N}}{q!} \sum \frac{1}{(2!)^{p_2} \dots (N!)^{p_N} p_1! \dots p_N!}, \\ P_n(q, Q) &= \frac{(N/n)^N n! q!}{N! (q + n - N)!} P_N(q, Q). \end{aligned} \right\} \quad (18)$$

Since the summation over all combinations of $p_1, \dots, p_N$, that are compatible with the given values of q and Q is the most laborious part, this shows that the simultaneous distribution of q and Q under the second design is rather easily found if that under the first design is known. Since

$$G = \frac{n \sum x_i^2}{N} - N = \frac{nQ}{N} - N, \quad (19)$$

the distribution of G follows directly from that of Q which in turn is found in the obvious way from that of q and Q . The greater part of the work was done with punched cards by the Mechanical Operation Section of the Institute. The exact conditional distributions of Fisher's criterion were found to be as shown in Table 4.

Table 4

G	$P(G)$	G	$P(G)$	G	$P(G)$	G	$P(G)$
<i>Design: $N = 16, n = 16$</i>							
0	—	16	0.13914	32	0.00332	48	0.00009
2	0.00014	18	0.10746	34	0.00276	50	0.00006
4	0.00310	20	0.06998	36	0.00157	52	0.00002
6	0.02302	22	0.04922	38	0.00094	54	0.00001
8	0.07212	24	0.02740	40	0.00044	56	0.00001
10	0.12829	26	0.01959	42	0.00020	58	0.00001
12	0.17017	28	0.01172	44	0.00023	60	0.00001
14	0.16260	30	0.00621	46	0.00016	62 +	0.00001
							1.00000

<i>Design: $N = 16, n = 32$</i>							
16	0.01040	44	0.04089	72	0.00024		
20	0.07343	48	0.02256	76	0.00010		
24	0.18562	52	0.00936	80	0.00010		
28	0.23396	56	0.00468	84	0.00006		
32	0.19905	60	0.00246	88	0.00002		
36	0.13970	64	0.00126	92	0.00001		
40	0.07549	68	0.00060	96 +	0.00001		
					1.00000		

<i>Design: $N = 16, n = 80$</i>							
64	0.20039	114	0.01344	164	0.00008		
74	0.36995	124	0.00427	174	0.00004		
84	0.25504	134	0.00193	184	0.00001		
94	0.10990	144	0.00051	194 +	0.00001		
104	0.04430	154	0.00013				
					1.00000		

The fact that the sum of the probabilities is exactly one in all three designs is a coincidence, since no adjustment to that effect was made. For the approximative distributions the normal approximation (4) was not used, since four of the values of ν were too small. K. Pearson's (1922) *Tables of the Incomplete Γ -Function* were used instead. These give cumulative probabilities $I(u, p)$, starting from the left, of a variate u , associated with a parameter p , which are related to the ν and the χ^2 of equations (9) by

$$p = \frac{1}{2}\nu - 1, \quad u = \chi^2/(2\nu)^{\frac{1}{2}}.$$

For this purpose, using the correction for continuity, equations (11) and (12) were replaced by

$$p = \frac{N(n-1)}{2(N-1)} - 1, \quad u = \frac{NG - n}{\sqrt{[2N(n-1)(N-1)]}}. \quad (20)$$

Application of (20) gave the results shown in Table 5.

Table 5

Design	$N = 10, n = 10$	$N = 16, n = 8$	$N = 16, n = 16$	$N = 16, n = 32$	$N = 16, n = 80$
p	4	2.733	7	15.533	41.133
u	$0.24846(G-1)$	$0.13802(2G-1)$	$0.18856(G-1)$	$0.13129(G-2)$	$0.08216(G-5)$

The comparisons of the cumulative distributions, starting from high values of G , are shown in Table 6.

Comparing the two designs with $N = n$, it appears that the approximation improves moderately, if N increases from 10 to 16. In the latter design, the 1 and 5% levels of significance are correct. In the former design the 5% level is correct. The approximate 1% level is 22, whereas the exact probability of G being equal to 22 or greater is 0.014. The error is hardly of practical importance. It may be safely concluded that for designs with $N = n$, the approximation is good enough if $N = 16$ or greater. Comparing the four designs with $N = 16$, it appears that the approximation deteriorates if n increases from 8 to 80 and the cell-expectation decreases from 2 to 0.2. Yet even in the last design the nominal 1 and 5% levels are still correct, namely at $G = 124$ and 114 respectively.

In the last design there are 80 cells, each with expectation 0.2. We have seen before that the approximation is excellent in a design with 100 cells with expectation 0.067, 100 cells with expectation 0.13, 100 cells with expectation 0.93 and 100 cells with expectation 1.87. It seems that the disturbing influence of small expectations is effectively counteracted by large numbers of cells, especially if there are also considerable numbers of cells with somewhat greater expectations.

5. APPLICATION OF THEORY TO NUMERICAL EXAMPLES

Example 1

The records of Farmer & Chambers (1939) of the accidents of 166 London bus-drivers over a period of 5 years constitute a design with 5 columns and 165 rows, since one driver, who had not a single accident, had to be omitted. So $m = 5$, $n = 165$. The yearly totals and the grand total are:

y_1	y_2	y_3	y_4	y_5	N
301	256	268	250	255	1330

Table 6. *Cumulative distributions, starting from high values of G.*

G	Exact	Approx.	G	Exact	Approx.	G	Exact	Approx.	G	Exact	Approx.
<i>Design: N = 10, n = 10 (exact distribution taken from Vessereau)</i>											
2	1.000	1.000	10	0.433	0.402	18	0.039	0.042	26	0.003	0.003
4	0.983	0.972	12	0.240	0.270	20	0.024	0.020	28	0.002	0.001
6	0.869	0.852	14	0.157	0.154	22	0.014	0.010	30	0.001	0.000
8	0.657	0.651	16	0.071	0.082	24	0.006	0.004	32	0.000	—
<i>Design: N = 16, n = 8 (exact distribution taken from Vessereau)</i>											
2	1.000	1.000	9	0.384	0.379	16	0.038	0.039	23	0.004	0.003
3	0.989	0.986	10	0.274	0.288	17	0.027	0.027	24	0.003	0.002
4	0.935	0.935	11	0.209	0.214	18	0.020	0.018	25	0.002	0.001
5	0.859	0.846	12	0.151	0.159	19	0.014	0.013	26	0.001	0.001
6	0.732	0.731	13	0.112	0.127	20	0.009	0.008	27	0.001	0.000
7	0.618	0.607	14	0.073	0.079	21	0.006	0.006	28	0.000	—
8	0.479	0.487	15	0.056	0.056	22	0.004	0.004			
<i>Design: N = 16, n = 16</i>											
0	1.000	1.000	12	0.603	0.609	24	0.047	0.047	36	0.002	0.001
2	1.000	0.999	14	0.441	0.453	26	0.028	0.025	38	0.001	0.000
4	0.997	0.994	16	0.301	0.316	28	0.016	0.014	40	0.001	—
6	0.974	0.963	18	0.194	0.208	30	0.010	0.007	42	0.001	—
8	0.902	0.887	20	0.124	0.131	32	0.007	0.004	44	0.000	—
10	0.773	0.762	22	0.075	0.079	34	0.004	0.002			
<i>Design: N = 16, n = 32</i>											
16	1.000	0.997	32	0.497	0.518	48	0.041	0.036	64	0.002	0.001
20	0.990	0.973	36	0.298	0.322	52	0.019	0.014	68	0.001	0.000
24	0.916	0.890	40	0.158	0.173	56	0.010	0.005	72	0.001	—
28	0.731	0.730	44	0.082	0.084	60	0.005	0.002	76	0.000	—
<i>Design: N = 16, n = 80</i>											
64	1.000	0.966	94	0.175	0.201	124	0.007	0.002	154	0.000	0.000
74	0.800	0.789	104	0.065	0.059	134	0.003	0.000			
84	0.430	0.480	114	0.020	0.012	144	0.001	0.000			

The numbers n_z of the personal totals z are:

z	n_z	z	n_z	z	n_z	z	n_z
1	2	6	17	11	9	16	6
2	3	7	14	12	6	17	3
3	14	8	14	13	2	19	1
4	17	9	12	14	6	21	3
5	21	10	13	15	1	32	1
Total							165

The sums of the inverses of row- and column-totals are:

$\sum z_i^{-1} = 28.5788, \quad \sum y_j^{-1} = 0.018881.$

The results of equations (7) and (8) are:

$\rho = 143.762, \quad \mu = 3.9880,$

$\sigma = 8.121, \quad \tau = 0.000084,$

$E(G) = 656.5, \quad \text{var } G = 1085.2.$

Using (5), the parameters c and ν are: $c = 1.210, \nu = 794.4.$

To illustrate the method of computation, recommended in § 3, for designs with more than two columns, the personal scores of the $n_z = 6$ drivers with $z = 12$ accidents in the total period were as follows (first scheme, step 3):

	x_1	x_2	x_3	x_4	x_5	Total
	3	2	4	1	2	12
	3	2	4	2	1	12
	2	3	1	2	4	12
	5	2	1	3	1	12
	2	2	2	2	4	12
	1	3	2	3	3	12
$s_i = \sum x_i:$	16	14	14	13	15	72
$q_i = \sum x_i^2:$	52	34	42	31	47	206

Continuing with step 3 of the first scheme, we shall have 20 rows (corresponding to 20 different z -values) with $2m + 5 = 15$ column headings. An extract from the resulting table is given in Table 7, showing the column 'totals' and the figures in that one of the 20 rows which corresponds to $z = 12$.

Note that only the 'totals' of columns 3, 4, 5, 6, 7, 8 and 9 are real totals, the other ones being product-sums, as described under step 4. There is no 'total' corresponding to column 1. With these results, χ^2 is computed as follows (step 5):

$$\sum x_{ij}^2 y_j^{-1} z_i^{-1} = \frac{100.425}{301} + \frac{73.909}{256} + \frac{80.047}{268} + \frac{73.708}{250} + \frac{73.428}{255}$$

$$= 1.5038,$$

$$G = 1330 \times 0.5038 = 670.1 \quad (\text{see equation (2)}),$$

$$\chi^2 = 1.210 \times 670.1 = 810.8 \quad (\text{see equation (5)}),$$

$$u = \sqrt{1621.6} - \sqrt{1587.8} = 0.42 \quad (\text{see equation (6)}),$$

so that the test does not reveal a significant departure from homogeneity.

Table 7

Column ... Heading ...	1 (z)	2 (z ⁻¹)	3 (n _z)	4 (s ₁)	5 (s ₂)	6 (s ₃)	7 (s ₄)	8 (s ₅)	9 (s)
Row (z = 12)	⋮ 12 ⋮	⋮ 0.0833 ⋮	⋮ 6 ⋮	⋮ 16 ⋮	⋮ 14 ⋮	⋮ 14 ⋮	⋮ 13 ⋮	⋮ 15 ⋮	⋮ 72 ⋮
'Total'	—	28.5788	165	301	256	268	250	255	1330

Column ... Heading ...	10 (q ₁)	11 (q ₂)	12 (q ₃)	13 (q ₄)	14 (q ₅)	15 (q)
Row (z = 12)	⋮ 52 ⋮	⋮ 34 ⋮	⋮ 42 ⋮	⋮ 31 ⋮	⋮ 47 ⋮	⋮ 206 ⋮
'Total'	100.425	73.909	80.047	73.708	73.428	401.517

Example 2

As an illustration of the outline for a design with two columns, consider Adelstein's data of accidents of 122 shunt-workers, specified for two periods of 5 and 6 years, cited from Arbous & Kerrich (1951). The complete records are given in the form shown in Table 8 (see step 1 of the second scheme of § 3).

In agreement with the scheme, the number of accidents in the 5-year period is denoted by x and that of the 6-year period by $z - x$. If the difference between the periods makes this worth while, it saves some computation to take the shortest one for x . From the 8 diagonals of this concise table the lines of the 5-column table are easily written down (step 2 of the scheme for $m = 2$) (Table 9).

Table 8. *Adelstein's accident data.*

	$z-x$							
x	0	1	2	3	4	5	6	Total
0	21	18	8	2	1	.	.	50
1	13	14	10	1	4	1	.	43
2	4	5	4	2	1	.	1	17
3	2	1	3	2	.	1	.	9
4	.	.	1	1	.	.	.	2
5	.	.	.	.	.	.	.	0
6	.	.	.	.	.	.	.	0
7	.	1	.	.	.	.	.	1
Total	40	39	26	8	6	2	1	122

Table 9

(z)	(z^{-1})	(n_z)	(x)	(x^2)
1	1.0000	31	13	13
2	0.5000	26	22	30
3	0.3333	19	26	48
4	0.2500	7	12	26
5	0.2000	9	17	39
6	0.1667	5	13	39
7	0.1429	1	4	16
8	0.1250	3	12	62
274	54.984	101	119	74.836

In agreement with step 3 of the scheme, only the 'totals' of the columns headed (n_z) and (x) are sums, the other ones being product-sums. Hence

$m = 2, \quad n = 101, \quad N = 274, \quad y_1 = 119, \quad y_2 = 155, \quad \Sigma y_j^{-1} = 0.014855, \quad \Sigma z_i^{-1} = 54.984,$

$G = 274 \times (74.836 \times 0.014855 - 119/155) = 94.26 \quad (\text{step 4}),$

$\rho = 63.370, \quad \mu = 0.99634,$

$\sigma = 17.885, \quad \tau = 0.00026,$

$E(G) = 100.37, \quad \text{var } G = 92.99 \quad (\text{see equations (7) and (8)}),$

$c = 1.079, \quad \nu = 108.3, \quad \chi^2 = 101.7, \quad u = -0.42 \quad (\text{see equations (5) and (6)}).$

Again there is no significant departure from homogeneity.

Example 3

We have applied the same test to the absences of the 248 shift-workers recorded by Arbous & Sichel (1954) and again there was no significance. The personal totals of these shift-workers will be used here to illustrate the test of fit of an assumed negative binomial distribution (case (iii) of § 2). We denote the total number of workers by n , the numbers of workers with x absences in the total period by n_x and the expectations of these numbers by $\hat{n}_x = np_x$. The figures given in Table 10 are taken from Arbous & Sichel, table 1:

Table 10. Data for absences from Arbous & Sichel.

x	n_x	$\hat{n}_x$	$\hat{n}_x^{-1}$	$n_x^2/\hat{n}_x$	x	n_x	$\hat{n}_x$	$\hat{n}_x^{-1}$	$n_x^2/\hat{n}_x$
0	7	11.898	0.09	4.118	25	2	1.668	0.60	2.398
1	16	16.086	0.06	15.914	26	3	1.452	0.69	6.198
2	23	17.733	0.06	29.831	27	3	1.265	0.79	7.115
3	20	18.072	0.06	22.134	28	1	1.101	0.91	0.908
4	23	17.667	0.06	29.943	29	2	0.985	1.04	4.175
5	24	16.829	0.06	34.227	30	—	0.833	1.20	—
6	12	15.751	0.06	9.142	31	1	0.724	1.38	1.381
7	13	14.553	0.07	11.613	32	1	0.629	1.59	1.590
8	9	13.320	0.08	6.081	33	1	0.546	1.83	1.832
9	9	12.089	0.08	6.695	34	—	0.474	2.11	—
10	8	10.920	0.09	5.861	35	1	0.411	2.43	2.433
11	10	9.808	0.10	10.195	36	—	0.356	2.81	—
12	8	8.772	0.11	7.296	37	—	0.308	3.25	—
13	7	7.817	0.13	6.268	38	—	0.267	3.75	—
14	2	6.944	0.14	0.576	39	—	0.231	4.33	—
15	12	6.153	0.16	23.404	40	—	0.200	5.00	—
16	3	5.439	0.19	1.655	41	—	0.173	5.78	—
17	5	4.798	0.21	5.211	42	—	0.150	6.69	—
18	4	4.224	0.24	3.788	43	—	0.129	7.69	—
19	2	3.713	0.27	1.077	44	—	0.112	8.96	—
20	2	3.259	0.31	1.227	45	—	0.096	10.37	—
21	5	2.857	0.35	8.750	46	—	0.083	12.01	—
22	5	2.501	0.40	9.996	47	—	0.072	13.90	—
23	2	2.188	0.46	1.828	48	1	0.062	16.12	16.116
24	1	1.911	0.52	0.523	49+	—	0.386	2.59	—
Total					248	248	248.015	122.18	301.499

According to equations (10) we have:

$$G = 301.499 - 248 = 53.5, \quad E(G) = 49,$$

$$\text{var } G = 2 \times 49 - (2500 + 100 - 2)/248 + 122.18 = 209.7,$$

or, with the correction suggested under case (iii) of § 2, for the two estimated parameters of the distribution:

$$E(G) = 47 \times 49/49 = 47, \quad \text{var } G = 47 \times 209.7/49 = 201.$$

It follows by the equations (5) that

$$c = 2 \times 47/201 = 0.468, \quad \nu = 0.468 \times 47 = 22.0, \quad \chi^2 = 0.468 \times 53.5 = 25.0,$$

so that $P = 0.3$.

There is no significant departure from the negative binomial distribution, although full account is taken of the fact that one worker had 48 absences. The computations would be less if all expectations less than five or so were grouped, but the power of the test would be heavily reduced. Judging from the results of § 4, we suppose that the approximation is still usable, although the inevitable grouping is restricted to expectations smaller than 0.06.

Example 4

As an illustration of the testing of an assumed Poisson distribution, according to the cases (iii) and (iv) of § 2, we have selected the accidents of the shunt-workers in the 5-year subperiod (see Example 2 above), because Arbous & Kerrich (1951, p. 356) asserted that for each subperiod 'the distribution of accidents was found to agree satisfactorily with the Poisson distribution':

Data		Case (iv)		Case (iii)		
x	n_x	xn_x	x^2n_x	$\hat{n}_x$	$\hat{n}_x^{-1}$	$n_x^2/\hat{n}_x$
0	50	0	0	46.00	0.0	54.3
1	43	43	43	44.87	0.0	41.2
2	17	34	68	21.88	0.0	13.2
3	9	27	81	7.11	0.1	11.4
4	2	8	32	1.73	0.6	2.3
5	0	0	0	0.338	3.0	0.0
6	0	0	0	0.055		
7	1	7	49	0.0087	15.7	15.7
	122	119	273	121.99	19.5	138.1

Beginning with the test of case (iv), we find according to equations (12):

$$\chi^2 = (122 \times 272 - 119^2)/118 = 161.2, \quad \nu = 119 \times 121/118 = 122.0.$$

It follows from (6) that

$$u = \sqrt{322.4} - \sqrt{243} = 2.37, \quad \text{giving } P = 0.009.$$

Comparing the columns headed n_x and $\hat{n}_x$, it will be seen that the significance is almost wholly due to the worker who had seven accidents. The Poisson probability of one or more workers having seven or more accidents is roughly equal to the P found with case (iv). Continuing with case (iii), it is necessary to curtail the distribution at a final class of $x = 6$ and more, with an expectation of 0.064. Using equations (10) it is found that

$$G = 138.1 - 122 = 16.1, \quad E(G) = 6, \quad \text{var } G = 12 - (49 + 14 - 2)/122 + 19.5 = 31.0,$$

or with the correction suggested under case (iii) of § 2,

$$E(G) = 5, \quad \text{var } G = 25.8,$$

and by equations (5),

$$c = 10/25.8 = 0.387, \quad \nu = 0.387 \times 5 = 1.94, \quad \chi^2 = 0.387 \times 16.1 = 6.23,$$

which is just significant at the 5 % level.

Application of the classical test requires a further curtailment at a final class of $x = 4$ and more with expectation 2.1. The result is $\chi^2 = 2.43$, $\nu = 3$, which is not even significant at the 30 % level. The strong evidence of the worker with seven accidents is completely destroyed if only the information that he had four or more accidents can be used.

Example 5

As an illustration of the application of the test of case (v) of § 2 to an assumed even distribution of points on an interval, we take the 108 explosions in British mines, in a period of 26,263 days, recorded by Maguire, Pearson & Wynn (1952). The period is divided into 108 subperiods of 243.176 days each:

x	n_x	xn_x	x^2n_x	$\hat{n}_x$
0	44	0	0	39.7
1	37	37	37	39.7
2	17	34	68	19.9
3	5	15	45	6.6
4	3	12	48	} 2.1
5	2	10	50	
	108	108	248	108.0

According to equations (13) and (6) we have:

$$\chi^2 = 108 \times (248 - 109)/107 = 140.3, \quad \nu = 108, \quad u = \sqrt{280.6} - \sqrt{215} = 2.09,$$

giving $P = 0.02$. The approximation is certainly good for $N = n = 108$, since according to § 4 it is already good for $N = n = 16$.

Again, the last column gives the Poisson expected group frequencies $\hat{n}_x$, for a mean of 1.00. The standard χ^2 test gives $\chi^2 = 5.46$, $\nu = 3$, which is not significant at the 10 % level.

6. IMPORTANCE OF TESTS OF HOMOGENEITY IN STATISTICS OF ABSENTEEISM

With the rising standards of life and social security the importance of psychological and sociological grounds of human absence behaviour seems to increase in relation to purely medical reasons. The urgent need of a rational control of absenteeism calls for more knowledge of this complex phenomenon. If statistics are to be of any service for this difficult problem, some statistical model must be accepted first. The following is an extension of the

model invented by Greenwood & Woods (1919) when dealing with industrial accidents. For further reference see the extensive review of Arbous & Kerrich (1951).

Imagine that every worker has a private urn, with many white and few black balls, that he is frequently drawing a ball with replacement and that he has an accident, or starts an absence, if he draws a black ball. The proportions of black balls in the urns may be different. They are associated with the personal liability for accidents or absences of the workers concerned and with the risk and stress of their work. The same could be said of the speed of drawing balls. However, it would make no difference if the number of black balls of a certain worker were halved and his speed of drawing were simultaneously doubled. So we are free to associate the proportions of black balls and the speeds of drawing to different groups of causes. Let a worker's proportion of black balls be associated with his personal characteristics and his place and function in the concern, things that are relatively constant in time. Let his speed of drawing be associated with factors that are relatively variable in time and may roughly be divided into the following groups:

A. *Homogeneous factors*, that are the same for all workers of the observed population, such as weather conditions, very contagious diseases, economic activity, etc.

B. *Inhomogeneous factors*, that may be different for different groups of workers, such as moderately contagious diseases, changes in production scheme, in personnel policy, change of officers who are sociologically important, development of attitudes towards the concern, measures to improve working conditions and so on.

Since Greenwood first discussed the matter, attention has been mainly focused on the personal liability of workers. It was in order to isolate this factor, that most of the studies, reviewed by Arbous & Kerrich, were on carefully selected, homogeneous populations, place and function being the same for all workers and relations being such that the inhomogeneous factors (B) could hardly be suspected of being different for different groups of workers. As a result of those studies, it seems to have been established that the liability is different for different workers and that we may regard the proportion of black balls, multiplied by a suitable constant, as having approximately a gamma-distribution in such homogeneous populations. For more cautious assertions, see Arbous & Kerrich (1951). The practical consequences of these results were somewhat disappointing.

In this context it seems only natural to pay more attention to the inhomogeneous factors (B), since these seem most liable to practical control. If one or more factors, which influence absenteeism, are varying in time, the speed of drawing will be multiplied by a number that is also varying in time. If those factors are changing similarly for all workers, the multiplier will also change similarly for all workers, that is, it will be equal for all workers at any fixed moment. Without loss of generality we may consider that the drawings occur simultaneously in this case. So, in any subperiod, every worker would have the same number of drawings and the distribution of black balls in the $m \times n$ -design with $(m-1)(n-1)$ degrees of freedom would be homogeneous in the statistical sense. Conversely, if the test gives an insignificant result, it may be concluded that there is no reason to investigate the factors (B). In the opposite case one will conclude that a partial epidemic has occurred during the period of observation, that is, some of the workers have changed their absence behaviour in relation to the others. Several statistical methods might help to find out which of the factors (B) is responsible for that change. For example, if there are only two subperiods, one might single out a group of workers who had most of their absences in the first subperiod and another group of those who had them mainly in the second. Then one

has to consider in which other respects these two groups are different. At the present stage of our knowledge it will often be difficult to find a reasonable explanation for an apparent statistical inhomogeneity. A reliable diagnostic system can be built up only if a large number of tentative diagnoses is already available.

REFERENCES

- ARBOUS, A. G. & KERRICH, J. E. (1951). Accident statistics and the concept of accident-proneness. *Biometrics*, **7**, 340-432.
- ARBOUS, A. G. & SICHEL, H. S. (1954). New techniques for the analysis of absenteeism data. *Biometrika*, **41**, 77-90.
- DAWSON, R. B. (1954). A simplified expression for the variance of the χ^2 -function on a contingency table. *Biometrika*, **41**, 280.
- FARMER, E. & CHAMBERS, E. G. (1939). A study of accident proneness among motor drivers. *Rep. Industr. Hlth Res. Bd*, no. 84.
- FISHER, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.
- GREENWOOD, M. & WOODS, H. M. (1919). The incidence of industrial accidents upon individuals with special reference to multiple accidents. *Rep. Industr. Hlth Res. Bd*, no. 4.
- GRÜNEBERG, H. & HALDANE, J. B. S. (1937). Tests of goodness of fit applied to records of Mendelian segregation in mice. *Biometrika*, **29**, 144-53.
- HALDANE, J. B. S. (1937). The exact value of the moments of the distribution of χ^2 , used as a test of goodness of fit, when expectations are small. *Biometrika*, **29**, 133-43.
- HALDANE, J. B. S. (1939). The mean and variance of χ^2 , used as a test of homogeneity, when expectations are small. *Biometrika*, **31**, 346-55.
- MAGUIRE, B. A., PEARSON, E. S. & WYNN, A. H. A. (1952). The time intervals between industrial accidents. *Biometrika*, **39**, 168-80.
- PEARSON, K. (1922). *Tables of the Incomplete Γ -Function*. Cambridge University Press.
- PEARSON, K. (1932). Experimental discussion of the (χ^2, P) test for goodness of fit. *Biometrika*, **24**, 351-81.
- VESSEREAU, A. (1958). Sur les conditions d'application du critérium χ^2 de Pearson. *Bullet. de l'Institut. Intern. de Statist.* Tome 36, 3e Livr., 87-101. Stockholm.

JUMP ANALYSIS

BY G. H. JOWETT AND WENDY M. WRIGHT

Department of Statistics, University of Sheffield

1. INTRODUCTION

When a length of time series is such as to fall naturally into k successive sections, each consisting of n evenly spaced successive terms, it is often desired to carry out statistical tests to determine whether the probability structure of the series is unaffected by the transition from one section to the next or suffers a change in mean. If variation within sections is random, the conventional analysis relevant to this issue would be an analysis of variance between and within sections, the observed variance of section means being compared, using the F -test, with their hypothetical variance, estimated from the pooled variance within sections, under the null hypothesis of a common underlying population mean. In many applications, however, the structural assumption of random variation about a mean varying from section to section is replaced by that of sections of an autocorrelated series subject to different displacements of mean, potentially due to assignable causes, from section to section; when this is so, the analysis described above is no longer valid, and a modified or substitute form of analysis must be found. Unfortunately the analysis of variance pattern of test leads to two difficulties.

The first of these arises because the observed means of different sections do not vary independently under the null hypothesis, with the result that the variance of section means cannot be taken as a chi-squared variate. This is not an insuperable difficulty, because at worst the expected value and standard error of this mean square can be worked out from the correlational properties of the series if these can be sufficiently well determined, thus providing a test of significance which is asymptotically valid.

The second and more fundamental difficulty arises from the fact that the between sections mean square is essentially the mean semi-squared difference between all possible pairs of section means, and its expectation can only be determined if the mean semi-squared difference between any such pair expected under the null hypothesis can be predicted. Unfortunately this prediction always involves correlational properties of the series with lags greater than the length of a section, even if the means of the pair are those of adjacent sections, since terms separated by more than a section length are involved; these properties can only be inferred from the within section comparison provided by the data when it can safely be assumed that the autocorrelation dies away within the length of a section, or in the unusual event of extensive extrapolation being possible. To overcome this difficulty it is necessary to abandon the analysis of variance approach altogether, and find an alternative.

The test proposed in this paper uses what will be called the *between-section jump statistic* B_α , defined as the mean semi-squared difference (or jump) between the means of the last α ($\leq \frac{1}{2}n$) terms of each section and the first α terms of the following section.

In § 2 there will be developed an asymptotic test using B_α as a criterion which is appropriate under certain generally applicable working assumptions of smoothness; in § 3 this test will be given more detailed examination, particularly as regards the choice of an α for maximal sensitivity, under the more restrictive but still widely applicable assumption of

locally Markoff variation (defined and discussed in Davies & Jowett (1958)). Under favourable conditions, and for suitable choice of α , it will be shown that the statistic

$$J_\alpha \equiv \ln (B_\alpha / W_\alpha), \quad (1)$$

(where W_α is the mean semi-squared difference of all comparable within-section jumps) has expectation zero and approximate asymptotic standard error given by

$$\sigma_{J_\alpha} \doteq \sqrt{\left(\frac{2 \cdot 2}{k-1} \right)} \quad (2)$$

under the null hypothesis, and provides the test of significance required.

2. SAMPLING VARIANCE OF B_α IN TERMS OF SERIAL VARIATION PARAMETERS

Denote the series by $x(t)$, the kn observations being given by $t = 1, 2, \dots, kn$ and having structure

$$x[(s-1)n+u] = \mu_s + \xi(s-1n+u), \quad (3)$$

where $\xi(t)$ will be taken as a stationary series for purposes of exposition, though this assumption may be relaxed and the techniques developed in this paper applied to the more general class of smoothly heteromorphic series (following the principle established in Jowett (1957)), where only the local variational properties of $\xi(t)$ are effectively stationary.

The second-order variational properties of $\xi(t)$ are defined by the serial variation function $\delta(\tau)$, given by the formula

$$\delta(\tau) \equiv E\{\frac{1}{2}[\xi(t) - \xi(t+\tau)]^2\}, \quad (4)$$

a symmetric function of τ related to the more commonly used variance σ^2 and autocorrelation function $\rho(\tau)$ by the equation

$$\delta(\tau) = \sigma^2(1 - \rho(\tau)). \quad (5)$$

Denote the average of a number α of successive values of $\xi(t)$ centred at t , called a *centralized average*, by $\bar{\xi}_\alpha(t)$; thus, for example, $\bar{\xi}_2(1\frac{1}{2})$ is the mean of $\xi(1)$ and $\xi(2)$. For successive values of t , $\bar{\xi}_\alpha(t)$ is itself a stationary series, with second-order variational properties defined by its serial variation function

$$\delta_{\bar{\xi}_\alpha}(\tau) \equiv E\{\frac{1}{2}[\bar{\xi}_\alpha(t) - \bar{\xi}_\alpha(t+\tau)]^2\}, \quad (6)$$

which for convenience in printing will be denoted by $\delta_\alpha(\tau)$. It is related to $\delta(\tau)$ as follows:

$$\delta_\alpha(\tau) = \frac{1}{2\alpha^2} \sum_{m=1}^{\alpha} \sum_{m'=1}^{\alpha} E[\xi(t - \frac{1}{2} - \frac{1}{2}\alpha + m) - \xi(t - \frac{1}{2} - \frac{1}{2}\alpha + \tau + m)] \times [\xi(t - \frac{1}{2} - \frac{1}{2}\alpha + m') - \xi(t - \frac{1}{2} - \frac{1}{2}\alpha + \tau + m')] \quad (7)$$

$$= \frac{1}{2\alpha^2} \sum_{m=1}^{\alpha} \sum_{m'=1}^{\alpha} E\{-\frac{1}{2}[\xi(t - \frac{1}{2} - \frac{1}{2}\alpha + m) - \xi(t - \frac{1}{2} - \frac{1}{2}\alpha + m')]^2 - \frac{1}{2}[\xi(t - \frac{1}{2} - \frac{1}{2}\alpha + \tau + m) - \xi(t - \frac{1}{2} - \frac{1}{2}\alpha + \tau + m')]^2 + \frac{1}{2}[\xi(t - \frac{1}{2} - \frac{1}{2}\alpha + \tau + m) - \xi(t - \frac{1}{2} - \frac{1}{2}\alpha + m')]^2 + \frac{1}{2}[\xi(t - \frac{1}{2} - \frac{1}{2}\alpha + m) - \xi(t - \frac{1}{2} - \frac{1}{2}\alpha + \tau + m')]^2\} \quad (8)$$

$$= \frac{1}{2\alpha^2} \sum_{m=1}^{\alpha} \sum_{m'=1}^{\alpha} [\delta(m-m'-\tau) - 2\delta(m-m') + \delta(m-m'+\tau)], \quad (9)$$

$$= \frac{1}{2\alpha} \sum_{r=-(\alpha-1)}^{+(\alpha-1)} \left(1 - \frac{|r|}{\alpha}\right) \delta(\tau+r) - \frac{1}{2\alpha} \sum_{r=-(\alpha-1)}^{+(\alpha-1)} \left(1 - \frac{|r|}{\alpha}\right) \delta(r).$$

We may interpret $\delta_x(\tau)$ as a repeated centralized average of $\delta(\tau)$ adjusted by a constant to make $\delta_x(0)$ identically zero. In most practical applications $\delta(\tau)$ has a cusp at $\tau = 0$ and rises monotonically, for some distance at least; $\delta_x(\tau)$ behaves somewhat similarly, except that the cusp and curvature near the cusp are somewhat less pronounced. In other words, $\delta_x(\tau)$ is a smoothed and translated version of $\delta(\tau)$. To evaluate (9) exactly it is clearly necessary to know $\delta(\tau)$ up to the lag $\tau + \alpha - 1$, the values round τ being particularly critical.

The expectation, under the null hypothesis, of the semi-squared difference between the means of the s th and s' th section is

$$\delta_n(\overline{s-s'}n), \quad (10)$$

but the serial variation statistics

$$d(\tau) \equiv \text{Av } \frac{1}{2}[x(t) - x(t+\tau)]^2 \quad (11)$$

required to estimate $\delta(\tau)$ may be computed from within-section differences only up to $\tau = n - 1$. It is thus evident why such expectations cannot in general be estimated and why the analysis of variance test based on them cannot in general be applied.

The between-section jump statistic B_x is given by

$$B_x = \frac{1}{k-1} \sum_{s=1}^{k-1} \frac{1}{2} (\bar{x}_x(sn + \frac{1}{2} + \frac{1}{2}\alpha) - \bar{x}_x(sn + \frac{1}{2} - \frac{1}{2}\alpha))^2. \quad (12)$$

Under the null hypothesis that μ_s is constant for $s = 1, \dots, k$, we have

$$E(B_x) = \delta_x(\alpha). \quad (13)$$

Now for a general normal stationary stochastic series $y(t)$ it may be shown (cf. Jowett, 1955a) that

$$\text{cov} [\frac{1}{2}(y(t) - y(t+\tau))^2, \frac{1}{2}(y(t') - y(t'+\tau))^2] = \frac{1}{2}[\Delta''_\tau \delta_y(t-t')]^2, \quad (14)$$

where Δ''_τ is the second central difference operator at interval τ .

Applying this formula when $y \equiv \xi_x$, under the null hypothesis of constant μ_s we have

$$\text{var } B_x = \frac{1}{(k-1)^2} \sum_{s=1}^k \sum_{s'=1}^k \frac{1}{2} [\Delta''_\alpha \delta_x(\overline{s-s'}n)]^2 \quad (15)$$

$$= \frac{1}{k-1} \sum_{m=-(k-2)}^{+(k-2)} \frac{1}{2} \left(1 - \frac{|m|}{k-1}\right) [\Delta''_\alpha \delta_x(mn)]^2 \quad (16)$$

$$= \frac{1}{k-1} \left(2[\delta_x(\alpha)]^2 + \sum_{m=1}^{k-2} \phi_m \right), \quad \text{say}, \quad (17)$$

where

$$\Phi_m = \left(1 - \frac{|m|}{k-1}\right) [\Delta''_\alpha \delta_x(mn)]^2 \quad (18)$$

$$= \left(1 - \frac{|m|}{k-1}\right) \left[\sum_{r=-(\alpha-1)}^{+(\alpha-1)} \frac{1}{\alpha} \left(1 - \frac{|r|}{\alpha}\right) \Delta''_\alpha \delta(mn+r) \right]^2, \quad (19)$$

the form (19) being obtained from (18) by substituting (9), the second term of which is annihilated by the differencing operation.

The first term in the bracket in (17) may be calculated from values of $\delta(\tau)$ which go only as far as $\tau = 2\alpha - 1$, to estimate which serial variation statistics are available from within-section comparisons provided that $\alpha \leq \frac{1}{2}n$. No direct estimates of $\delta(mn+r)$ (or equivalently, $\delta_x(mn)$) are available for $|m| > 1$, but in principle at least the second differences in (18) may be estimated by averages of all possible products of within-section differences at lag $mn+r$, since if

$$p_x(t, \tau) \equiv [\bar{x}_x(t - \frac{1}{2}\tau - \frac{1}{2}\alpha) - \bar{x}_x(t - \frac{1}{2}\tau + \frac{1}{2}\alpha)] [\bar{x}_x(t + \frac{1}{2}\tau - \frac{1}{2}\alpha) - \bar{x}_x(t + \frac{1}{2}\tau + \frac{1}{2}\alpha)], \quad (20)$$

$$E\{p_x(t, \tau)\} = \Delta''_\alpha \delta_x(\tau), \quad (21)$$

whenever $t, t + \alpha$ come from the same section and $t + \tau, t + \tau + \alpha$ come from the same section. The difficulties attached to this procedure arise through the necessity for graduating these averages by a suitable formula (cf. Bartlett, 1946) before substituting in (13).

It is also useful to have a formula for the asymptotic logarithmic variance of B_α :

$$\text{var}(\log_e B_\alpha) \sim \frac{1}{k-1} \left(2 + \sum_{m=1}^{k-2} \Psi_m \right), \quad (22)$$

where
$$\Psi_m = \frac{\Phi_m}{[\delta_\alpha(\alpha)]^2}, \quad (23)$$

since the essentially positive nature of B_α , coupled with some evidence from sampling experiments, suggest that $\log B_\alpha$ will have a sampling distribution closer to normal than that of B_α . The null hypothesis may thus be tested by comparing

$$\log(B_\alpha/\delta_\alpha(\alpha)), \quad (24)$$

with a suitable multiple of its standard error, which is the square root of (22).

Since the variance of the location parameters μ_s is estimated by subtracting from an observed value B_α its expectation under the null hypothesis, given by (13), the asymptotically optimal value of α from the point of view of power is that which proves to give the smallest value for (17). In the general case a special *ad hoc* investigation is called for, though consideration of the limiting cases of strictly random variation within sections at one extreme and smooth continuous variation at the other suggests (as confirmed in §3 below under the more restrictive assumption of locally Markoff variation) that smaller values of α are likely to be optimal for smoother series, and vice versa.

The general technique just described is both laborious and difficult to apply in practice; two forms of simplification are possible. The first, discussed here, applies if $\delta(\tau)$ satisfies certain smoothness assumptions; the second, discussed in §3, applies if $\delta(\tau)$ may be represented approximately within a suitable interval centred at $\tau = 0$, by a function of the form

$$\theta(1 - \rho^{|\tau|}). \quad (25)$$

The assumptions are less restrictive in the first case, but it is necessary to fit the serial variation function $\delta(\tau)$, whereas in the second case this is avoided.

The statistic B_α is a member of a class of statistics called *local statistics*, the general sampling properties of which have been discussed in Jowett (1955*a*). It is a function of local comparisons, being a mean square of differences involving adjacent groups of terms of the series, and therefore its sampling formulae are dominated by terms involving $\delta(\tau)$ for values of τ which are small compared with the total time interval covered by the data. The necessary condition on $\delta(\tau)$ is that it shall become increasingly linear as τ increases, in the sense described in Jowett (1955*a*), with the result that $\delta_\alpha(\tau)$ also becomes increasingly linear and $\Delta''_\alpha \delta_\alpha(\tau)$ converges rapidly to zero (increasingly so the smaller the value of α) as $|\tau|$ increases. If this condition is satisfied the terms Φ_m of (17) make contributions which decrease rapidly as m increases. The whole of (17) will be dominated by the first term in the bracket, and the size of this will furnish a good guide for the optimization of the choice of α . Having provisionally chosen α in this way, the practical statistician may reasonably follow the course of neglecting all but possibly the first one or two of $\Phi_1, \Phi_2, \dots, \Phi_{k-2}$, unless there is positive evidence to warrant either their retention or a reduction in α sufficient to justify their omission.

Such evidence of a rather subjective kind may be furnished by scrutiny of the variogram constructed from the statistics $d(1), \dots, d(n-1)$, which may give some indication of rough upper bounds for the second differences $\Delta''_{\alpha}\delta(n-\alpha+1)$, $\Delta''_{\alpha}\delta(n+\alpha-1)$, which straddle the range of second differences $\Delta''_{\alpha}\delta(\tau)$, from which $\Delta''_{\alpha}\delta_{\alpha}(n)$ is computed as a weighted average. More objective evidence would be the significance of a relevant mean of the product $p_{\alpha}(t, mn)$ of (20); this is essentially a straightforward mean of terms from a time series, albeit generated from $\xi(t)$ in a rather complicated way, and its standard error may be estimated by the general method given in Jowett (1955*b*). Neither method is particularly sensitive, but at least they provide some opportunity to check the working assumptions.

Usually advantage may be taken of the availability of a direct estimate of the dominant term of (17) which does not require explicit consideration of the serial variation curve. Corresponding to the statistic B_{α} , which is the sum of semi-squares of what might be described as between-section jumps, is a statistic W_{α} which is the corresponding sum of semi-squares of comparable within-section jumps, defined by

$$W_{\alpha} = \frac{1}{k(n-2\alpha+1)} \sum_{s=1}^k \sum_{u=\alpha}^{n-\alpha} \frac{1}{2} (\bar{x}_{\alpha}(s-1, n + \frac{1}{2} + u + \frac{1}{2}\alpha) - \bar{x}_{\alpha}(s-1, n + \frac{1}{2} + u - \frac{1}{2}\alpha))^2. \quad (26)$$

Independently of the truth of the null hypothesis,

$$E(W_{\alpha}) = \delta_{\alpha}(\alpha), \quad (27)$$

so that W_{α} provides an estimate both of the expectation of B_{α} under the null hypothesis and of the dominant term in (17). On general grounds the relative error of this estimate may be expected to increase with α , being the same as that of B_{α} for $\alpha = \frac{1}{2}n$, and this must be borne in mind in using it in (17) as a basis for a large-sample test of significance. More explicit consideration of the error of W_{α} under the assumption of locally Markoff variation will be given in § 3.

Although the exposition above is strictly applicable only under the assumptions that $\xi(t)$ is stationary and normal, the test is robust under departures from these assumptions in certain directions. It was shown in Jowett (1957) that formulae for sampling properties of local statistics in stationary series could be applied as an approximation in non-stationary series having the properties of smooth heteromorphy and acting normality. For a full discussion the reader is referred to the paper in question, but the implications in the present case are that formulae such as (17) may be applied when the variation in $\xi(t)$ is locally homogeneous and normal within intervals comparable in length to the maximum lag τ for which the terms of the formula are non-negligible. The rapid convergence of (17) is vulnerable mainly to short-term periodic tendencies of autoregressive type, unless α can be taken as a multiple of the quasi-periods.

3. THE JUMP CRITERION J_{α} AND ITS APPLICATION IN LOCALLY MARKOFF SERIES

If α is small enough for the error in W_{α} to be neglected, and for all except the dominant term in (17) to be neglected, it becomes possible to carry out the test of significance for B_{α} by calculating the jump criterion J_{α} defined by (1) and testing it against a null hypothesis of value of zero, using its asymptotic sampling variance as given by (22), which reduces to $2/(k-1)$. In this form the test itself becomes extremely simply to carry out, the difficulty being rather one of assessing whether the conditions necessary for its validity are satisfied

Also, considerations of power may dictate a value of α larger than that for which they can be satisfied.

To explore the potentialities of J_α as a test statistic, a numerical investigation was carried out using the two parameter Markoff serial variation function:

$$\delta(\tau) = \theta(1 - \rho^{|\tau|}). \quad (28)$$

Numerical values taken were $\rho = 0.0(0.1)0.9, 0.95$; $n = 12$; $k = \infty$; $\alpha = 1(1)6$. It was unnecessary to take numerical values for θ , which as a scale parameter does not affect the sampling properties of J_α .

From (16), for large k

$$\text{var } B_\alpha \sim \frac{1}{k} \sum_{m=-\infty}^{+\infty} \frac{1}{2} [\Delta''_\alpha \delta_\alpha(mn)]^2. \quad (29)$$

$$\text{Similarly, } \text{var } W_\alpha = \frac{1}{k^2(n-2\alpha+1)^2} \sum_{s=1}^k \sum_{s'=1}^k \sum_{u=\alpha}^{n-\alpha} \sum_{u'=\alpha}^{n-\alpha} \frac{1}{2} [\Delta''_\alpha \delta_\alpha(\overline{s-s'}n - u + u')]^2, \quad (30)$$

$$\sim \frac{1}{k(n-2\alpha+1)^2} \sum_{m=-\infty}^{+\infty} \sum_{u=\alpha}^{n-\alpha} \sum_{u'=\alpha}^{n-\alpha} \frac{1}{2} [\Delta''_\alpha \delta_\alpha(mn - u + u')]^2, \quad (31)$$

$$\text{cov}(B_\alpha, W_\alpha) = \frac{1}{k(k-1)(n-2\alpha)} \sum_{s=1}^{k-1} \sum_{s'=1}^k \sum_{u=\alpha}^{n-\alpha} \frac{1}{2} [\Delta''_\alpha \delta_\alpha(\overline{s-s'}n + u)]^2, \quad (32)$$

$$\sim \frac{1}{k(n-2\alpha+1)} \sum_{m=-\infty}^{+\infty} \sum_{u=\alpha}^{n-\alpha} \frac{1}{2} [\Delta''_\alpha \delta_\alpha(mn + u)]^2. \quad (33)$$

Formulae (29), (31), (33) were computed in the cases described, taking $\theta = 1$, and the results substituted in the formulae

$$\text{var}(B_\alpha - W_\alpha) = \text{var } B_\alpha - 2 \text{cov}(B_\alpha, W_\alpha) + \text{var } W_\alpha, \quad (34)$$

$$\text{var } J_\alpha \sim \frac{\text{var}(B_\alpha - W_\alpha)}{[E(W_\alpha)]^2}, \quad (35)$$

to give the results in Table 1.

The following broad conclusions emerge.

(i) Below $\rho = 0.6$, $k \text{var}(B_\alpha - W_\alpha)$ decreases to a minimum at or near $\alpha = 6$ (i.e. $\frac{1}{2}n$); above $\rho = 0.6$, it increases from a minimum for $\alpha = 1$ (i.e. the smallest possible value of α). This indicates which values of α give the most powerful test.

(ii) The pattern of rise or fall in $k \text{var}(B_\alpha - W_\alpha)$ with α strongly resembles that of $E(W_\alpha)$, except for $\rho = 0.6$, where variation in either is small. This confirms the appropriateness in this case of the practice recommended in § 2, namely, choosing as the optimal α the value minimizing $E(W_\alpha)$.

(iii) Below $\rho = 0.6$, $k \text{var } J_\alpha < 2.2$ when $\alpha \leq 5$ (actually when $\alpha \leq 6$ for $\rho < 0.3$); at and above $\rho = 0.6$, $k \text{var } J \leq 2.2$ when $\alpha = 1$. Thus if we are prepared to content ourselves with a slightly suboptimal value of α in certain cases, we can cover ourselves against errors of the first kind at least to the usual extent by working with the formula (2) which provides us with a simple test along the lines described at the beginning of the section which is not restricted to small values of α in cases where a large value of α would be optimal. The factor 2.2 is a convenient round number which is not unacceptably large.

(iv) In the cases mentioned in (iii), the error committed by neglecting all but the dominant term in (17) proves to be negligible. Evidently the increase of the factor from 2 to 2.2 is occasioned by the need to allow for the sampling error of W_α .

Table 1. *Variances and covariances of jump statistics for Markoff serial variation function with $n = 12$*

ρ α	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
<i>k var B_α</i>											
1	2.000	1.620	1.280	0.980	0.720	0.500	0.320	0.180	0.080	0.020	0.005
2	0.500	0.540	0.558	0.548	0.508	0.440	0.346	0.237	0.127	0.038 <i>1</i>	0.010
3	0.222	0.268	0.315	0.357	0.385	0.389	0.359	0.290	0.183 <i>1</i>	0.065 <i>1</i>	0.019 <i>1</i>
4	0.125	0.160	0.200	0.246	0.292	0.329	0.344	0.316 <i>1</i>	0.233 <i>2</i>	0.096 <i>3</i>	0.031 <i>3</i>
5	0.080	0.106	0.138	0.178	0.225	0.274	0.315 <i>1</i>	0.326 <i>2</i>	0.274 <i>4</i>	0.130 <i>6</i>	0.044 <i>4</i>
6	0.056	0.075	0.100	0.134	0.177 <i>1</i>	0.230 <i>1</i>	0.286 <i>3</i>	0.329 <i>6</i>	0.313 <i>10</i>	0.167 <i>10</i>	0.059 <i>7</i>
(Percentage increase over dominant term alone shown in italics)											
<i>k var W_α</i>											
1	0.265	0.202	0.152	0.111	0.078	0.052	0.032	0.018	0.008	0.002	0.001
2	0.088	0.096	0.100	0.100	0.094	0.082	0.066	0.046	0.025	0.008	0.002
3	0.058	0.074	0.089	0.104	0.116	0.120	0.114	0.095	0.063	0.023	0.007
4	0.045	0.061	0.080	0.102	0.126	0.148	0.161	0.157	0.124	0.055	0.018
5	0.047	0.065	0.089	0.119	0.156	0.197	0.235	0.253	0.222	0.103	0.038
6	0.056	0.075	0.100	0.134	0.177	0.230	0.286	0.329	0.313	0.167	0.059
<i>k var ($B_\alpha - W_\alpha$)</i>											
1	2.083	1.701	1.354	1.043	0.770	0.537	0.345	0.195	0.087	0.022	0.006
2	0.519	0.563	0.586	0.581	0.545	0.477	0.381	0.265	0.145	0.045	0.012
3	0.231	0.281	0.331	0.378	0.412	0.423	0.400	0.333	0.221	0.083	0.025
4	0.123	0.158	0.199	0.246	0.294	0.338	0.368	0.363	0.299	0.140	0.047
5	0.074	0.098	0.128	0.168	0.218	0.281	0.353	0.418	0.414	0.218	0.078
6	0.056	0.077	0.108	0.151	0.212	0.298	0.412	0.536	0.574	0.325	0.113
<i>E(W_α)</i>											
1	1.000	0.900	0.800	0.700	0.600	0.500	0.400	0.300	0.200	0.100	0.050
2	0.500	0.520	0.528	0.523	0.504	0.469	0.416	0.344	0.252	0.138	0.072
3	0.333	0.366	0.397	0.422	0.439	0.441	0.424	0.380	0.302	0.179	0.097
4	0.250	0.282	0.317	0.351	0.382	0.406	0.414	0.396	0.338	0.216	0.122
5	0.200	0.230	0.262	0.298	0.335	0.370	0.396	0.400	0.363	0.248	0.145
6	0.167	0.193	0.224	0.258	0.297	0.337	0.373	0.395	0.378	0.275	0.167
<i>E(W_α)/E(W_1)</i>											
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.500	0.578	0.660	0.747	0.840	0.938	1.040	1.147	1.260	1.380	1.440
3	0.333	0.407	0.496	0.603	0.732	0.882	1.060	1.267	1.510	1.790	1.940
4	0.250	0.313	0.396	0.501	0.637	0.812	1.035	1.320	1.690	2.160	2.440
5	0.200	0.256	0.328	0.426	0.558	0.740	0.990	1.333	1.815	2.480	2.900
6	0.167	0.214	0.280	0.369	0.495	0.674	0.933	1.317	1.890	2.750	3.340
<i>k var J_α</i>											
1	2.08	2.10	2.12	2.13	2.14	2.15	2.16	2.16	2.17	2.18	2.20
2	2.07	2.09	2.10	2.12	2.14	2.17	2.20	2.24	2.29	2.35	2.37
3	2.08	2.10	2.10	2.12	2.14	2.17	2.22	2.30	2.43	2.60	2.70
4	1.97	1.98	1.99	2.00	2.02	2.06	2.14	2.31	2.61	3.00	3.17
5	1.84	1.85	1.86	1.89	1.94	2.05	2.25	2.61	3.14	3.55	3.73
6	2.00	2.06	2.15	2.26	2.41	2.63	2.96	3.44	4.02	4.30	4.07

The numerical investigation just described involved a considerable amount of computation and yet covers only one particular case. To justify application to other cases it was necessary to examine the relevance and robustness of the conclusions drawn under the following conditions: (a) departure of the serial variation function from strict Markoff form; (b) departure of n from the selected value $n = 12$; (c) the value of k not large.

(a) If the serial variation function may be represented in the form

$$\delta(\tau) = \theta(1 - \rho^{|\tau|}) + \omega(\tau), \quad (36)$$

where $\omega(\tau)$ is negligible for $\tau < 2\alpha - 1$ and has a second derivative so small that $\Delta_\alpha'' \omega(\tau)$ is negligible for $\tau > \alpha$, the effect of the term $\omega(\tau)$ on (21), (23) and (24) will simply be the addition of an average of these negligible quantities. This more general form of serial variation function, described as *locally Markoff* and discussed in a wider context in Davies & Jowett (1958), appears to have sufficient extra flexibility to serve as a robust working assumption in the absence of quasi-cycle variation, analogously with the normal distribution in classical theory.

For example, the function

$$\delta(\tau) = 0.81(1 - 0.8^{|\tau|}) + 0.19(1 - 0.2^{|\tau|}) \quad (37)$$

illustrates a common type of divergence from the strict Markoff form, namely a slower decline to zero in the initially steep gradient, and is roughly representable in the form (36) for $\rho = 0.6$. The sampling parameters of Table 1 computed for this function proved to be similar in essentials in their pattern of variation, α being nearly optimal at 1, $E(W)$ varying little, and $k \text{ var } J$ being less than 2.2 for $\alpha = 1$ (and also $\alpha = 2, 3$).

The robustness of the assumption of locally Markoff serial variation is of course greater the smaller the value of α ; however, its use need not be restricted to small values of α .

(b) Supplementary investigations similar to that yielding Table 1, taking $n = 6$, $\alpha = 1, 2, 3$, $\rho = (0.2)^2, (0.6)^2$ and $(0.8)^2$ and $n = 24$, $\alpha = 1, 8, 10, 11, 12$, $\rho = \sqrt{0.2}, \sqrt{0.6}, \sqrt{0.8}$ were carried out to test the extension of conclusions (i)–(iv) to smaller and larger values of n .

The values of ρ taken are equivalent to 0.2, 0.6, 0.8 in terms of section length.

It was found that, as before, the patterns of $k \text{ var } (B_\alpha - W_\alpha)$ and $E(W_\alpha)$ were similar; α was optimal at unity for the highest values of ρ and not greatly suboptimal at one less than its maximum for the lowest value of ρ .

There was also evidence to suggest that the transitional value of the Markoff parameter in terms of section length increases with n , the middle value being below the transition for $n = 6$, above it for $n = 24$. Also 2.2 remained a safe approximation to $k \text{ var } J_\alpha$ in the corresponding cases.

(c) As a rough check on the applicability of the asymptotic theory for values of k not large and to gain experience, jump analyses were carried out for twenty artificially constructed Markoff series in each of the cases:

$$\begin{aligned} \rho = 0.8; \quad n = 6, k = 10; \quad \alpha = 1. \\ \rho = 0.36; \quad n = 6, k = 10; \quad \alpha = 2. \\ \rho = 0.2; \quad n = 6, k = 10; \quad \alpha = 3. \end{aligned}$$

It was found that the distribution of J_α was consistent with symmetry, satisfying a χ^2 goodness of fit test for a normal curve, and that only two out of the sixty values fell above the quasi-normal upper 5% point at 0.813 ($= 1.645 \sqrt{[2.2/9]}$).

The formula (2) is therefore proposed as the usual criterion for the significance of J_α on the grounds that it makes a small allowance which is adequate to cover the non-vanishing terms of the summation in (16) and the error of W_α at least in the appropriate Markoff cases, and which in the absence of evidence to the contrary, may reasonably be taken to do so in other cases, or at least will be better than making no allowance at all.

4. COMPARISON WITH THE ANALYSIS OF VARIANCE TEST

Since only adjacent sections are contrasted in jump analysis, one would expect some loss of power compared with a method based on the same principle as the analysis of variance, where the means of all possible pairs of sections are effectively compared. The latter approach is, of course, not usually possible when the serial variation curve is not known fairly completely; nevertheless, the comparison is still of some interest, and we shall draw it for large k and an assumed random normal variation in the parameters μ_s characterized by a standard deviation σ_μ , taking the serial variation curve as being unit Markoff in form, and taking $n = 12$ as in §3. Under these conditions, writing $B_{\alpha 0}$ as the value of B_α under the null hypothesis, we have

$$(k-1)B_\alpha = (k-1)B_{\alpha 0} + \sum_{s=1}^{k-1} \frac{1}{2}(\mu_s - \mu_{s-1})^2 + \sum_{s=1}^{k-1} [\bar{\xi}_\alpha(sn + \frac{1}{2} + \frac{1}{2}\alpha) - \bar{\xi}_\alpha(sn + \frac{1}{2} - \frac{1}{2}\alpha)](\mu_s - \mu_{s-1}). \quad (38)$$

The covariances of the three terms on the right-hand side of (38) are clearly zero; the variance of the first is given by (17), and the variances of the second and third are easily shown to be given respectively by

$$\sigma_\mu^4(3k-4) \quad (39)$$

and

$$2(k-2)\sigma_\mu^2[\Delta''_\alpha \delta_\alpha(0) - \Delta''_\alpha \delta_\alpha(n)] + 2\sigma_\mu^2 \delta_\alpha(n), \quad (40)$$

so that asymptotically

$$k \text{ var } B_\alpha \sim k \text{ var } B_{\alpha 0} + 3\sigma_\mu^4 + 2\sigma_\mu^2[\Delta''_\alpha \delta_\alpha(0) - \Delta''_\alpha \delta_\alpha(n)]. \quad (41)$$

Also

$$E(B_\alpha) = E(B_{\alpha 0}) + \sigma_\mu^2, \quad (42)$$

so that the estimate of σ_μ^2 given by

$$\text{Est}_{J_\alpha} \sigma_\mu^2 = B_\alpha - E(B_{\alpha 0}) \equiv B_\alpha - E(W_\alpha) \quad (43)$$

has the same sampling variance as B_α .

Alternatively, we have an estimate from the calculated variance $\text{var}_c \bar{x}_n$ of section means given by

$$\text{Est}_v \sigma_\mu^2 = \text{var}_c \bar{x}_n - E[\text{var}_c \bar{\xi}_n]. \quad (44)$$

Hence, similarly to the preceding argument,

$$\begin{aligned} \text{var}(\text{Est}_v \sigma_\mu^2) &= \text{var}(\text{var}_c \bar{x}_n) \\ &= \text{var}(\text{var}_c \bar{\xi}_n) + \text{var}(\text{var}_c \mu) + 2 \text{var}[2 \text{cov}_c(\bar{\xi}_n, \mu)], \end{aligned} \quad (45)$$

where the suffix c implies that the variance or covariance is a calculated, not expected, one. Now, using formula (5) of Bartlett (1946) we have

$$\text{var}(\text{var}_c \bar{\xi}_n) \sim \frac{2}{k} \sum_{s=-\infty}^{+\infty} \{\delta_{\bar{x}_n}(\infty) - \delta_{\bar{x}_n}(sn)\}^2 \quad (46)$$

and it is a standard result that

$$\text{var}(\text{var}_c \mu) \sim \frac{2\sigma_\mu^4}{k}. \quad (47)$$

For the third term

$$\begin{aligned} \text{var}[2 \text{cov}_c(\bar{\xi}_n, \mu)] &\sim \frac{4}{k^2} E \left\{ \sum_{s, s'=1}^k \bar{\xi}_n((s-1)n + \frac{1}{2} + \frac{1}{2}n) \bar{\xi}_n((s'-1)n + \frac{1}{2} + \frac{1}{2}n) \mu_{s-1} \mu_{s'-1} \right\} \\ &\sim \frac{4}{k^2} \sum_{s=1}^k \sigma_\mu^2 \text{var}[\bar{\xi}_n((s-1)n + \frac{1}{2} + \frac{1}{2}n)] \\ &= \frac{4}{k^2} \sigma_\mu^2 (\delta_{\bar{x}_n}(\infty) - \delta_{\bar{x}_n}(0)). \end{aligned} \quad (48)$$

We have used these formulae for asymptotic variances to compute the efficiency of $\text{Est}_{J_\alpha} \sigma_\mu^2$ relative to $\text{Est}_v \sigma_\mu^2$ for $n = 12$ and a number of different values of α , ρ and σ_μ . The results are given in Table 2. They all settle down to a limit of 67%, i.e. $\frac{2}{3}$ as $\sigma_\mu \rightarrow \infty$.

Table 2. *Asymptotic efficiency of estimate of σ_μ^2 from B_α compared with that obtained from $\text{var}_c \bar{x}_n$*

σ_μ^2 (unit = σ_ξ^2)	$\rho = 0.0$ Opt. $\alpha = 6$	$\rho = 0.1$ Opt. $\alpha = 6$	$\rho = 0.6$ Opt. $\alpha = 1$	$\rho = 0.95$ Opt. $\alpha = 1$
$\frac{1}{4}$	54	54	63	1700
1	63	63	67	262
4	66	66	67	99
25	67	67	67	71
100	67	67	67	68

Table 3. *Increased sampling variance of estimate of σ_μ^2 from $\text{var}_c \bar{x}_n$ due to presence of within-section serial correlation*

Ratio of variance (percentage) to that for estimate from random data.

σ_μ^2 (unit = σ_ξ^2)	$\rho = 0.1$	$\rho = 0.6$	$\rho = 0.95$
$\frac{1}{4}$	110	254	1818
1	103	140	346
4	101	110	145
25	100	102	106
100	100	100	101

approaching it from above in the cases $\rho \geq 0.7$ and from below in the cases $\rho \leq 0.6$, the optimum jump-analysis value of α being chosen in each case. In the cases where the limit is approached from above, the relative efficiency can be considerably greater than 100% when σ_μ is of the same order as σ_ξ (which has the value of unity in our calculations) or smaller. It is in precisely these cases that W_α supplies a stable estimate of $E(B_{\alpha 0})$, so that this asymptotic comparison has a relevance to practical jump analysis; in the other cases it is more abstract, since W_α has a variance comparable with that of $B_{\alpha 0}$. However, the graphs are consistent with what we would expect on general grounds, namely, that as the variation of $\xi(t)$ approaches randomness, a test based on variance is more efficient than any jump

analysis. It should not, of course, be forgotten that even in cases where the acting Markoff serial variation curve has a small parameter, the true curve may diverge from it for lags greater than the section length, rendering a test based on variance impossible to carry out, for the reasons given in § 1.

That jump analysis can be more efficient than a test based on variance, were such a test practicable, in cases where the parameter is large is not intuitively surprising; a large parameter reflects smoothness, and discontinuities from section to section become increasingly obvious as this smoothness increases. This gain can be almost large enough to offset the loss due to the presence of serial correlation, as can be seen from comparison of Tables 2 and 3; this occurs when the entry in Table 2 is almost as large as the corresponding entry in Table 3. The information yielded by jump analysis on serially correlated data as a fraction of that which would be yielded by variances of random data with the same dispersion is given roughly by dividing the former by the latter, and is evidently sometimes quite small.

Illustrative example

The data in Table 4 below consist of 156 terms of an industrial time series which has been arbitrarily broken up into $k = 12$ successive sections of $n = 13$ consecutive terms each. The analysis given is fuller than would be warranted in many practical applications, but even so begs certain questions (e.g. testing the goodness of fit of the serial variation curve) which lie outside the scope of this paper and might reasonably form the subject of inquiry.

Table 4. *Weekly determinations of percentage of free carbon in coking coal as charged (origin = 46, unit = 0.1)*

Term	Section											
	1	2	3	4	5	6	7	8	9	10	11	12
1	40	71	67	61	39	46	58	35	78	46	24	38
2	27	47	69	49	27	65	51	50	32	38	32	32
3	40	65	58	65	46	60	38	43	43	21	35	25
4	45	44	58	59	53	62	57	54	43	26	35	27
5	34	65	70	59	45	50	53	41	52	24	28	19
6	42	53	48	58	25	59	64	45	56	38	44	26
7	30	56	53	58	35	46	69	34	46	30	46	28
8	45	61	56	44	34	50	58	40	44	42	36	22
9	54	58	59	49	29	53	71	43	35	37	33	41
10	50	96	66	45	37	72	52	50	40	19	31	34
11	56	72	72	45	48	56	43	35	35	44	37	26
12	57	86	66	51	55	56	27	21	35	39	28	29
13	49	47	68	46	36	57	23	50	49	25	32	33

The first necessary step is to calculate and compare two estimates of W_α from alternate sections for $\alpha = 1, \alpha = 6$. We have calculated it for these and for intermediate values of α , the bulk of the labour consisting of computing moving differences of moving averages of α terms for the 156 successive terms.

In view of the remarks in the paragraph following that where (16) occurs, a choice of $\alpha = 6$ seems optimal, giving the value 30.1 for W_x . The value of B_x for $\alpha = 6$ is 32.1, so that

$$J_\alpha = \ln(32.1/30.1) = 0.0643. \quad (49)$$

From (2) the *prima facie* standard error of J_α is

$$\sqrt{(2.2/11)} = 0.447, \quad (50)$$

so that J_α is not significant on the 5 % level. No further analysis is necessary, since any factor more correct than 2.2 in (50) would certainly not be less than 2. In more critical circumstances, however, the need for revision of (50) might be investigated along the following lines.

(a) *Scrutiny of the variogram*

For fitting a locally Markoff serial variation function to the data the least-squares method (Davies & Jowett, 1958) would be appropriate; however, the breaking of the series into short sections and the non-availability of $d(15)$ prevent the tables in this paper from being used, and circumstances do not justify the labour of an *ad hoc* computation. However, as was done by Davies & Jowett it may fairly be assumed that the fitted function will reproduce

Table 5. *Values of W_x for coal series*

α	Odd sections	Even sections	All sections
1	61.9	87.8	74.9
2	48.0	34.9	41.4
3	51.1	31.4	41.2
4	50.8	31.9	41.3
5	35.1	32.5	33.8
6	27.5	32.7	30.1

$d(1)$ (which is identical with W_1), and it is both convenient and reasonable to obtain the remaining equation necessary for the estimation of the two Markoff parameters in (28) by equating W_6/W_1 to $E(W_6)/E(W_1)$ in Table 1.

Hence from Table 5

$$\text{Est } \theta(1 - \text{Est } \rho) = 74.9,$$

and the value of 0.402 for W_6/W_1 suggests the value 0.33 for $\text{Est } \rho$ (using inverse linear interpolation in the appropriate section of Table 1), so that $\text{Est } \theta = 112$.

The variogram has been plotted in Fig. 1, this fitted curve being shown. The later serial variation statistics are very erratic, being based on only very few comparisons, and their departure from the fitted curve is probably of no consequence; the standard error of $d(2)$ (cf. Table 11 of Davies & Jowett, 1958) is roughly 16 and the discrepancy of 26, though not convincing in itself, taken along with the evidence of $d(3)$, $d(4)$ and $d(5)$ suggests a serial variation curve such as that shown by the broken line as being more likely *prima facie*. Following the suggestion in the antepenultimate paragraph of § 2, we extrapolate the curve to a value of 127 at $\tau = 14$, and find that $\Delta_6^a \delta(8) = -15$ for the broken curve as compared

with -12 for the Markoff curve; a fair working assumption is that $\Delta_6''\delta(18)$ is negligible for both curves. Taking the average suggests values of Ψ_1 in (22) of about $0.06, 0.04$ respectively, both rather trifling. Hence the use of (2) is not contra-indicated.

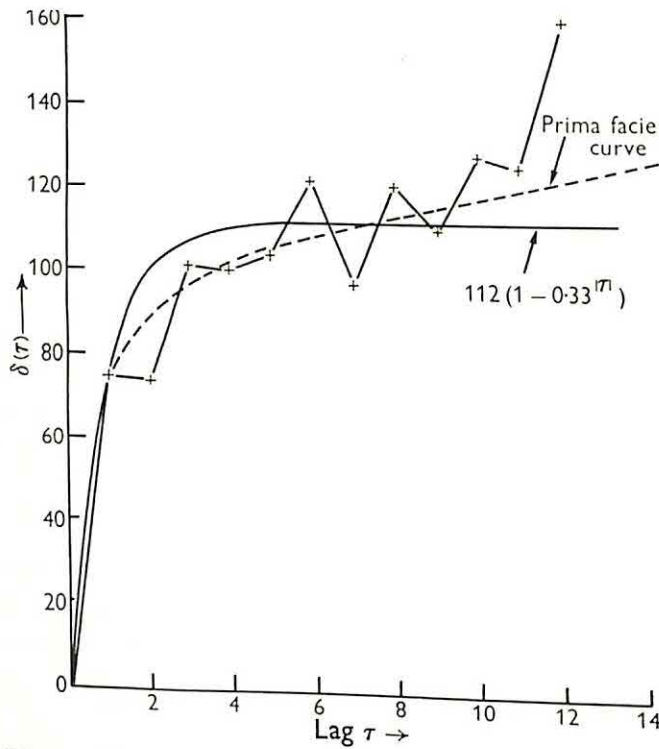


Fig. 1. Variogram and fitted curves for free carbon series.

(b) Assessment of mean product of differences

Using (20) we obtain the following series of values (rounded) for $p_6(t, 13)$ from the data:

t	$p(t)$	t	$p(t)$	t	$p(t)$
13	149	65	-1	117	-35
14	233	66	0	118	-27
26	5	78	0	130	7
27	78	79	-3	131	-19
39	-3	91	1	143	5
40	-59	92	45	144	-18
52	-5	104	86	$\bar{p}: 20$	
53	-15	105	26		

If we were to take the mean at its face value, it would suggest a value of 0.4 for Ψ_1 , giving

$$\text{var}(\log B_\alpha) \div \frac{1}{11} \left\{ 2 + \frac{10}{11} \times \left(\frac{20}{30} \right)^2 \right\} = \frac{2.4}{11},$$

suggesting a value slightly greater than that given in (50) for the standard error of J_α . However, the variation in the figures is considerable, and there is no *prima facie* case for relying on this correction.

It is possible to make only a rough estimate of the standard error of $\bar{p}$, in view of the nature and shortness of the series. The serial variation curve for $p_6(t, 13)$ may be expected to be complicated in shape, in view of the many common values of $x(t)$ involved in neighbouring values of $p(t)$, and its estimation from the data is quite impracticable. However, the series of means of the pairs of adjacent values is more tractable, a fair working assumption being that the serial variation curve flattens out by a lag of about 26, for which the successive means have no values of $x(t)$ in common.

The analysis of semi-squared differences (following Jowett, 1955a) between the eleven resulting means is as follows:

Source	Sum of s.s.d.'s	Number	Mean s.s.d.
All	433,450	121	3,582
Lag 0	0	11	0
Lags ± 1	38,143	20	1,907
Other lags	395,307	90	4,392

Hence

$$\text{Est } \sigma_{\bar{p}} = \sqrt{(4392 - 3582)} = 28.$$

This can only be a very rough estimate, as will be obvious to any statistician who looks at the figures on which it is based; however, it suggests that the value of $\bar{p}$ could easily be fortuitous, and there are no compelling grounds for making the correction suggested above.

We would like to thank Dr H. C. Hamaker, who first drew our attention to the need for a solution of this problem; to Miss H. M. Davies and Miss D. Kirk, who have assisted us with the computation; and to Mr J. Hebden of the National Coal Board and to the Carbonization Branch of the Board, who kindly supplied the data.

REFERENCES

- BARTLETT, M. S. (1946). On the theoretical specification and sampling of autocorrelated time series. *J. R. Statist. Soc. (Suppl.)*, **8**, 27-41.
- DAVIES, H. M. & JOWETT, G. H. (1958). The fitting of Markoff serial variation curves. *J. R. Statist. Soc. B*, **20**, 120-42.
- JOWETT, G. H. (1955a). Sampling properties of local statistics in stationary stochastic series. *Biometrika*, **42**, 160-9.
- JOWETT, G. H. (1955b). The comparison of means of sets of observations from sections of independent stochastic series. *J. R. Statist. Soc.* **17**, 208-27.
- JOWETT, G. H. (1957). Statistical analysis using local properties of smoothly heteromorphic stochastic series. *Biometrika*, **44**, 454-63.

THE RANDOM WALK [IN CONTINUOUS TIME] AND ITS APPLICATION TO THE THEORY OF QUEUES

BY C. R. HEATHCOTE AND J. E. MOYAL

Australian National University, Canberra

SUMMARY. The random walk in continuous time has been considered by a number of authors for particular types of boundary conditions, e.g. the unrestricted random walk by Irwin (1937); the random walk with one absorbing or one reflecting barrier by Ledermann & Reuter (1954), and subsequently by Bailey (1954). The present paper gives a unified theory of these processes which includes previous work on the subject as special cases and readily yields the solution for any other consistent form of boundary conditions. Several of these are of interest in the theory of queues with exponential service and arrival times. The random walk with reflecting barriers at O and N (§4) is the single server queue problem where the number of possible customers is limited to N (Morse, 1958). By letting $N \rightarrow \infty$ we obtain the familiar single server queue problem solved previously in the papers by Ledermann & Reuter and by Bailey referred to above and also by Clarke (1956), Champenowne (1956) and Conolly (1958).

The passage to a 'diffusion' process is also considered. The method given provides an alternative to the well-known method of images (Chandrasekhar, 1943; Bartlett, 1956). In §5 we consider a more general process where the transition intensities λ and μ depend linearly on the particle's position in a bounded interval, and are constant outside this interval. This enables us to give the Laplace transform of the probability generating function for the N server queue, a problem recently considered by Karlin & McGregor (1958).

1. INTRODUCTION

A particle will be said to describe a random walk in continuous time if it has constant chances λ, μ per unit time of taking one unit step respectively to the right or left on the real axis. For prescribed boundary conditions, such as absorbing or reflecting barriers at given positions, we ask for the distribution of the particle's position at time t after the start of the process. Specifically, we choose the origin on the real axis (as can be done without loss of generality) so that the initial position of the particle at $t = 0$ is k , an integer, and for each of the processes studied we seek $P_{k,n}(t)$, the probability of a transition from k to n in time t . We use the 'backward' equation (Bartlett, 1956; Feller, 1957)

$$\frac{dP_{k,n}(t)}{dt} = -(\lambda + \mu)P_{k,n}(t) + \lambda P_{k+1,n}(t) + \mu P_{k-1,n}(t). \quad (1.1)$$

The probability generating function $G_k(z, t) = \sum_n P_{k,n}(t) z^n$ then satisfies

$$\frac{\partial G_k(z, t)}{\partial t} = -\alpha G_k(z, t) + \lambda G_{k+1}(z, t) + \mu G_{k-1}(z, t), \quad (1.2)$$

where $\alpha = \lambda + \mu$. Clearly $G_k(z, 0) = z^k$. Introducing the Laplace transform

$$g_k(z, s) = \mathcal{L}[G_k(z, t)] = \int_0^\infty e^{-st} G_k(z, t) dt,$$

it follows from (1.2) that

$$(s + \alpha)g_k = \lambda g_{k+1} + \mu g_{k-1} + z^k \quad (\text{Re } s > 0). \quad (1.3)$$

The range of k depends on the problem under consideration. Thus for the unrestricted random walk $k = 0 \pm 1, \pm 2, \dots$; whereas for a barrier at the origin k is restricted to the non-negative integers.

The system consisting of the backward equation, together with the side conditions, has a unique 'honest' solution ('honest' in the sense that $\sum_n P_{k,n}(t) \equiv 1$, or equivalently $G_k(1, t) \equiv 1$, $g_k(1, s) = s^{-1}$); in fact 'honesty' implies uniqueness and conversely. This is well known to be the case for (1.1) without side conditions (cf. Bartlett, Feller, *loc. cit.*). With side conditions it can be proved to follow from the general theory of discontinuous Markov processes given in Moyal (1957) by transforming the system into an equivalent integral equation of the type considered in that paper; the explicit solution for the single absorbing boundary case is given as an example of the general theory in Moyal (1959). The method of obtaining explicit solutions developed in the present paper is however much less laborious, especially for the more complicated side conditions, but we still rely on the general theory for the fact that the 'honesty' of each solution thus obtained implies its uniqueness.

Asymptotic equilibrium distributions are known to exist for this type of process. Their explicit expression is most easily obtained directly from the Laplace transform solution of (1.2) by an extension of Abel's Theorem (Widder, 1946, ch. v), which yields

$$\lim_{s \rightarrow +0} s g_k(z, s) = \lim_{t \rightarrow \infty} G_k(z, t), \quad (1.4)$$

whenever the limit on the right-hand side exists. It is in fact sufficient for (1.4) to be true that the right-hand side exist as a Cesaro limit.

There is another type of asymptotic behaviour of the solutions which is of interest, namely, the passage to a 'diffusion' process. We effect this passage as follows: Let

$$\alpha = \lambda + \mu = \sigma^2/h^2, \quad \lambda - \mu = a/h, \quad x = nh, \quad y = kh. \quad (1.5)$$

We assume that each step is of magnitude h , that σ and a are constants, and we look for the distribution of the displacement x conditional on y as $h \rightarrow 0$. Let

$$f(x | y; t) = \lim_{h \rightarrow 0} P_{y/h, x/h}(t). \quad (1.6)$$

Carrying out this limiting procedure in (1.1) one arrives formally at the 'backward' diffusion equation with drift term

$$\frac{\partial}{\partial t} f(x | y; t) = \left\{ \frac{\sigma^2}{2} \frac{\partial^2}{\partial y^2} + a \frac{\partial}{\partial y} \right\} f(x | y; t). \quad (1.7)$$

It will be shown later that this passage to the limit in the solutions of the random walk problems yields the solutions of (1.7) with the appropriate boundary conditions.

2. THE GENERAL SOLUTION OF THE RANDOM WALK PROBLEM

To obtain the general solution of the difference equation (1.3) we need a particular solution and the solution of the homogeneous equation obtained by omitting the term z^k . Since the transition probabilities clearly remain invariant under translations of the origin it follows that $G_k(z, t) = z^k G_0(z, t)$ and hence

$$g_k(z, s) = z^k g_0(z, s) \quad (k = 0, \pm 1, \pm 2, \dots).$$

Substituting in (1.3) we have

$$g_0(z, s) = (s + \alpha - \lambda z - \mu z^{-1})^{-1} = \phi(z, s),$$

so that a particular solution of (1.3) is

$$g_k(z, s) = z^k \phi(z, s). \quad (2.1)$$

Substituting $g_k = u^k$ in the homogeneous equation we obtain

$$u^k(s + \alpha - \lambda u - \mu u^{-1}) = 0, \quad (2.2)$$

the two solutions of which are the zeros of $\phi(z, s)$, namely

$$u_1(s) = (2\lambda)^{-1} [s + \alpha - \sqrt{\{(s + \alpha)^2 - 4\lambda\mu\}}] \quad (2.3)$$

and

$$u_2(s) = (2\lambda)^{-1} [s + \alpha + \sqrt{\{(s + \alpha)^2 - 4\lambda\mu\}}].$$

The general solution of (1.2) is therefore

$$g_k(z, s) = z^k \phi(z, s) + A(z, s) u_1^k(s) + B(z, s) u_2^k(s). \quad (2.4)$$

We require that $g_k \rightarrow 0$ as $s \rightarrow \infty$, so, since $u_1 \rightarrow 0$ and $u_2 \rightarrow \infty$ as $s \rightarrow \infty$, A must be bounded, while B must tend to 0 with $1/s$ faster than u_2^{-k} . For problems involving only one boundary we set $B \equiv 0$ so that (2.4) reduces to

$$g_k(z, s) = z^k \phi(z, s) + A(z, s) u_1^k(s), \quad (2.5)$$

where A is determined by the boundary condition. Boundary conditions at two points determine both A and B .

If no boundary conditions are imposed we have the unrestricted random walk, the solution of which is (2.1). The Laplace transforms of the probabilities $P_{k,r}(t)$ may be readily obtained by expanding (2.1) in a Laurent series for values of z in an annulus centred on the origin, which contains the circle $|z| = 1$ for all s whose real part is positive. Clearly it is sufficient to do this for $k = 0$. We have then

$$g_0(z, s) = \lambda^{-1} (u_2 - u_1)^{-1} \left\{ \sum_{r=0}^{\infty} (u_1/z)^r + \sum_{r=1}^{\infty} (z/u_2)^r \right\}. \quad (2.6)$$

We note that $g_k(1, s) = 1/s$, so that the solution is unique. The inverse Laplace transforms may be found directly from tables (Erdelyi, 1954) to give for the probabilities

$$P_{0,r}(t) = \beta^r e^{-\alpha t} I_r(2t \sqrt{\lambda\mu}) \quad (r = 0, \pm 1, \pm 2, \dots), \quad (2.7)$$

where $\beta = (\lambda/\mu)^{1/2}$, and $I_r(x)$ is the modified Bessel function of the first kind. Throughout the rest of this paper the argument of Bessel functions will be suppressed, it being understood that it is $2t \sqrt{\lambda\mu}$. The mean and variance of the distribution (2.7) are easily found to be $(\lambda - \mu)t$ and $\alpha t = (\lambda + \mu)t$ respectively. Inverting (2.1) we have the generating function

$$G_k(z, t) = z^k \exp[-t(\alpha - \lambda z - \mu z^{-1})], \quad (2.8)$$

which may be expanded to verify the results obtained above.

If now we substitute from (1.5) and set $z = e^{i\theta h}$, $z^k = e^{i\theta y}$ in (2.4), then

$$\begin{aligned} w(\theta, s, y) &= \lim_{h \rightarrow 0} g_{y/h}(e^{i\theta h}, s) \\ &= e^{i\theta y} (s + \frac{1}{2}\sigma^2\theta^2 - i\alpha\theta)^{-1} + a(\theta, s) \exp\left\{-\frac{y}{\sigma} \sqrt{\left(2s + \frac{a^2}{\sigma^2} - \frac{ay}{\sigma^2}\right)}\right\} \\ &\quad + b(\theta, s) \exp\left\{\frac{y}{\sigma} \sqrt{\left(2s + \frac{a^2}{\sigma^2} - \frac{ay}{\sigma^2}\right)}\right\}, \end{aligned} \quad (2.9)$$

where $a(\theta, s) = \lim_{h \rightarrow 0} A(e^{i\theta h}, s)$ and $b(\theta, s) = \lim_{h \rightarrow 0} B(e^{i\theta h}, s)$ and $w(\theta, s, y)$ is the Laplace-Fourier transform of the solution of (1.7), i.e.

$$w(\theta, s, y) = \int_0^\infty e^{-st} dt \int_{-\infty}^\infty e^{i\theta x} f(x | y, t) dx. \quad (2.10)$$

In the unrestricted case $a = b = 0$ and

$$f(x | y; t) = (2\pi\sigma^2 t)^{-\frac{1}{2}} \exp \left\{ -\frac{(x - y - at)^2}{2\sigma^2 t} \right\}. \quad (2.11)$$

3. THE RANDOM WALK WITH ABSORBING BARRIERS

If the states O and N , say, are absorbing barriers, the process stops whenever the particle reaches either O or N so the boundary conditions are

$$g_0 = 1/s, \quad g_N = z^N/s. \quad (3.1)$$

The values of the constants A and B appropriate to this process are found by substituting the general solution (2.4) in (3.1); with these values of A and B (2.4) becomes

$$\begin{aligned} g_k &= (z^k - u_1^k) \phi + s^{-1} u_1^k + (s^{-1} - \phi) (z^N - u_1^N) (u_2^k - u_1^k) (u_2^N - u_1^N)^{-1} \\ &= (u_2^N - u_1^N)^{-1} \left\{ s^{-1} (u_1 u_2)^k (u_2^{N-k} - u_1^{N-k}) + \sum_{r=1}^k z^r [\lambda(u_2 - u_1)]^{-1} (u_2^r - u_1^r) (u_2^{N-k} - u_1^{N-k}) \right. \\ &\quad \left. \times (u_1 u_2)^{k-r} + \sum_{r=k+1}^{N-1} z^r [\lambda(u_2 - u_1)]^{-1} (u_2^k - u_1^k) (u_2^{N-r} - u_1^{N-r}) + z^N s^{-1} (u_2^k - u_1^k) \right\}. \end{aligned} \quad (3.2)$$

Inverting the Laplace transforms gives for the probabilities

$$P_{k,0}(t) = \beta^{-k} \sum_{j=0}^{\infty} \int_0^t \tau^{-1} e^{-\alpha\tau} \{ (2jN + k) I_{2jN+k} - [2(j+1)N - k] I_{2(j+1)N-k} \} d\tau, \quad (3.4)$$

$$P_{k,r}(t) = \beta^{r-k} \sum_{j=0}^{\infty} e^{-\alpha t} \{ I_{2(j+1)N-k+r} + I_{2jN+k-r} - I_{2(j+1)N-k-r} - I_{2jN+k+r} \} \quad (r = 1, 2, \dots, k), \quad (3.5)$$

$$P_{k,r}(t) = \beta^{r-k} \sum_{j=0}^{\infty} e^{-\alpha t} \{ I_{2(j+1)N+k-r} + I_{2jN-k+r} - I_{2(j+1)N-k-r} - I_{2jN+k+r} \} \quad (r = k+1, k+2, \dots, N-1), \quad (3.6)$$

$$P_{k,N}(t) = \beta^{N-k} \sum_{j=0}^{\infty} \int_0^t \tau^{-1} e^{-\alpha\tau} \{ [(2j+1)N - k] I_{2(j+1)N-k} - [(2j+1)N + k] I_{2(j+1)N+k} \} d\tau. \quad (3.7)$$

The absorption probabilities $P_{k,0}(t)$ and $P_{k,N}(t)$ are symmetrical in λ, μ and $k, N-k$ as expected. The expected value of the position r at time t conditional on initial position k is

$$\begin{aligned} E(r | k, t) &= k + (\lambda - \mu) t + (\lambda - \mu) \sum_{j=0}^{\infty} \int_0^t \tau^{-1} (t - \tau) e^{-\alpha\tau} \\ &\quad \times \{ \beta^{-k} \{ [2(j+1)N - k] I_{2(j+1)N-k} - (2jN + k) I_{2jN+k} \} + \beta^{N-k} \{ [(2j+1)N + k] \\ &\quad \times I_{2(j+1)N+k} - [(2j+1)N - k] I_{2(j+1)N-k} \} \} d\tau. \end{aligned} \quad (3.8)$$

The generating function of the stationary distribution obtained using (1.4) is

$$\begin{aligned} \lim_{t \rightarrow \infty} G_k(z, t) &= \lim_{s \rightarrow +0} sg_k(z, s) = \pi_k + z^N \gamma_k \\ &= \begin{cases} \frac{(\mu/\lambda)^k - (\mu/\lambda)^N}{1 - (\mu/\lambda)^N} + \frac{z^N \{1 - (\mu/\lambda)^k\}}{1 - (\mu/\lambda)^N} & \text{if } \lambda \neq \mu, \\ 1 - k/N + z^N k/N & \text{if } \lambda = \mu, \end{cases} \end{aligned} \quad (3.9)$$

which yields for the expectation $E(r | k, \infty)$

$$E(r | k, \infty) = \begin{cases} N\{1 - (\mu/\lambda)^k\} \{1 - (\mu/\lambda)^N\}^{-1} & \text{if } \lambda \neq \mu, \\ k & \text{if } \lambda = \mu. \end{cases} \quad (3.10)$$

The boundary value problem discussed above is the continuous time analogue of the classical 'gambler's ruin' problem, in which two gamblers with initial capital k and $N - k$ respectively compete. As may be deduced from general arguments the asymptotic results we derive are identical with those of Feller (1957, chapter xiv). The probabilities of ultimate ruin are π_k and γ_k respectively given by (3.9). Since $\pi_k + \gamma_k = 1$ ultimate absorption at either one of the boundaries is certain, and the expected lapse of time before this occurs is

$$\begin{aligned} E(t | k) &= \int_0^\infty t(P'_{k,0}(t) + P'_{k,N}(t)) dt \\ &= \begin{cases} \{k - N\{1 - (\mu/\lambda)^k\} \{1 - (\mu/\lambda)^N\}^{-1}\} (\mu - \lambda)^{-1} & \text{if } \lambda \neq \mu, \\ k(N - k) (2\lambda)^{-1} & \text{if } \lambda = \mu. \end{cases} \end{aligned} \quad (3.11)$$

The solution of the problem with a single absorbing barrier, say at the origin, may either be found from the above by letting $N \rightarrow \infty$ or directly using (2.5). Taking the limit as $N \rightarrow \infty$ in (3.2) we have for the generating function

$$g_k = (z^k - u_1^k) \phi + s^{-1} u_1^k. \quad (3.12)$$

The last term in the right-hand side of (3.2) tends to 0 as $N \rightarrow \infty$ since $u_1 < |z| < u_2$ and $u_1 < 1$, $u_2 > 1$ for $\text{Re } s > 0$. The probabilities in this case are

$$P_{k,0}(t) = k\beta^{-k} \int_0^t \tau^{-1} e^{-\alpha\tau} I_k d\tau, \quad (3.13)$$

$$P_{k,r}(t) = \beta^{r-k} e^{-\alpha t} \{I_{r-k} - I_{r+k}\} \quad (r = 1, 2, 3, \dots), \quad (3.14)$$

with expectation

$$E(r | k, t) = k + (\lambda - \mu)t - (\lambda - \mu)k\beta^{-k} \int_0^t \tau^{-1}(t - \tau) e^{-\alpha\tau} I_k d\tau. \quad (3.15)$$

These three equations yields incidentally the relations

$$\sum_{n=1}^{\infty} \beta^n \{I_{n-k} - I_{n+k}\} = \beta^k e^{\alpha t} - k \int_0^t \tau^{-1} e^{\alpha(t-\tau)} I_k d\tau \quad (3.16)$$

and, by putting $\lambda = \mu$,

$$\sum_{n=1}^{\infty} n \{I_{n-k}(\alpha t) - I_{n+k}(\alpha t)\} = k e^{\alpha t} \quad (k = 1, 2, \dots). \quad (3.17)$$

Asymptotic results are well known (see Feller, *loc. cit.*) and are

$$P_{k,0}(\infty) = \begin{cases} (\mu/\lambda)^k & \text{if } \lambda \geq \mu, \\ 1 & \text{if } \lambda \leq \mu. \end{cases} \quad (3.18)$$

$$P_{k,r}(\infty) = 0 \quad (r = 1, 2, \dots). \quad (3.19)$$

$$E(r | k, \infty) = \begin{cases} +\infty & \text{if } \lambda > \mu, \\ k & \text{if } \lambda = \mu, \\ 0 & \text{if } \lambda < \mu. \end{cases} \quad (3.20)$$

Probability densities of first passage times through a barrier are given by differentiating the appropriate absorption probability. For example, in the single barrier case, the probability density of the first passage time through the origin is, from (3.13),

$$\rho_k(t) dt = k\beta^{-k}t^{-1}e^{-\alpha t}I_k dt. \quad (3.21)$$

Passage to the 'diffusion process' limit in the case of a single absorbing barrier at O yields

$$w(\theta, s, y) = e^{i\theta y} (s - ia\theta + \frac{1}{2}\sigma^2\theta^2)^{-1} + [1/s - (s - ia\theta + \frac{1}{2}\sigma^2\theta^2)] \exp - \left\{ \frac{ay}{\sigma^2} + \frac{y}{\sigma} \sqrt{2s + \frac{a^2}{\sigma^2}} \right\}. \quad (3.22)$$

The second term in the right-hand side of (3.22) is the Laplace-Fourier transform of the probability of absorption. Inverting and differentiating with respect to t we obtain the probability density ρ of the first passage time at the origin

$$\rho(t, y) dt = \frac{y}{\sqrt{(2\pi\sigma^2 t^3)}} \exp \left\{ - \left(\frac{y^2}{2\sigma^2 t} + \frac{a^2 t}{2\sigma^2} + \frac{ay}{\sigma^2} \right) \right\} dt. \quad (3.23)$$

Note that
$$\int_0^\infty \rho(t, y) dt = \begin{cases} 1 & \text{if } a \leq 0, \\ e^{-2ay/\sigma^2} & \text{if } a \geq 0. \end{cases} \quad (3.24)$$

Let $f_0(x, t)$ be the density obtained by setting $y = 0$ in (2.11); then the probability density in the absence of absorption is

$$f(x | y, t) = f_0(x - y, t) - \int_0^t f_0(x, t - \tau) \rho(\tau, y) d\tau. \quad (3.25)$$

Set $Nh = L$, a constant. A more symmetric form of the transform $w(\theta, s, y)$ is obtained in the case of two absorbing barriers by translating the origin of x and y to $\frac{1}{2}L$,

$$w(\theta, s, y) = \psi e^{i\theta y} + \frac{(1/s - \psi) e^{-ay/\sigma^2}}{\sinh \omega L} \{ e^{(i\theta + a/\sigma^2)\frac{1}{2}L} \sinh \omega(y + \frac{1}{2}L) - e^{-(i\theta + a/\sigma^2)\frac{1}{2}L} \sinh \omega(y - \frac{1}{2}L) \}, \quad (3.26)$$

where
$$\psi = (s - ia\theta + \frac{1}{2}\sigma^2\theta^2)^{-1} \text{ and } \omega = \frac{1}{\sigma} \sqrt{2s + \frac{a^2}{\sigma^2}}.$$

The first passage time probability densities at $\frac{1}{2}L$ and $-\frac{1}{2}L$ are obtained by inverse transformation of the terms involving $e^{\pm i\theta L}$ in the right-hand side of (3.26). One finds

$$\left. \begin{aligned} \rho_+(t, y) dt &= \frac{\exp \left[-\frac{a}{\sigma^2} \{ (y - \frac{1}{2}L) + \frac{1}{2}at \} \right]}{\sqrt{(2\pi\sigma^2 t^3)}} \sum_{n=-\infty}^{\infty} [(2n + \frac{1}{2})L - y] \\ &\quad \times \exp \left[-\frac{1}{2\sigma^2 t} [(2n + \frac{1}{2})L - y]^2 \right] dt, \\ \rho_-(t, y) dt &= \frac{\exp \left[-\frac{a}{\sigma^2} \{ (y + \frac{1}{2}L) + \frac{1}{2}at \} \right]}{\sqrt{(2\pi\sigma^2 t^3)}} \sum_{n=-\infty}^{\infty} [(2n + \frac{1}{2})L + y] \\ &\quad \times \exp \left[-\frac{1}{2\sigma^2 t} [(2n + \frac{1}{2})L + y]^2 \right] dt, \end{aligned} \right\} \quad (3.27)$$

where ρ_+ and ρ_- are respectively the first passage time distributions at $\frac{1}{2}L$ and $-\frac{1}{2}L$. The probability density in the absence of absorption has an expression similar to (3.23),

$$f(x | y, t) = f_0(x - y, t) - \int_0^t f_0(x, t - \tau) [\rho_+(\tau, y) + \rho_-(\tau, y)] d\tau. \quad (3.28)$$

4. THE RANDOM WALK WITH REFLECTING BARRIERS

We now take the barriers at O and N to be reflecting ones, in which case the boundary conditions satisfied by (2.4) are

$$(s + \lambda)g_0 = \lambda g_1 + 1, \quad (s + \mu)g_N = \mu g_{N-1} + z^N. \quad (4.1)$$

Solving for the constants A and B as in §3 we obtain the solution

$$g_k = \phi\{X_k(z^N - z^{N+1}) + z^k + Y_k(1 - z^{-1})\}, \quad (4.2)$$

where

$$X_k = \lambda s^{-1}\{(u_2^{k+1} - u_1^{k+1}) - (u_1 u_2)(u_2^k - u_1^k)\}(u_2^{N+1} - u_1^{N+1})^{-1},$$

$$Y_k = \lambda s^{-1}(u_1 u_2)^{k+1}\{(u_2^{N+1-k} - u_1^{N+1-k}) - (u_2^{N-k} - u_1^{N-k})\}(u_2^{N+1} - u_1^{N+1})^{-1}.$$

(4.2) may be written as

$$g_k = \phi\{z^k + (1 - z^{-1})u_1^{k+1}(1 - u_1)^{-1}\} + \phi X_k\{z^N - z^{N+1} + (1 - z^{-1})u_1^{N+1}\}$$

$$\rightarrow \phi\{z^k + (1 - z^{-1})u_1^{k+1}(1 - u_1)^{-1}\} \quad \text{as } N \rightarrow \infty, \quad (4.3)$$

(4.3) is thus the solution for a single reflecting barrier at the origin, a result which may be checked by direct methods.

Expanding (4.2) and inverting the coefficient of z^r yields for the probabilities:

$$P_{k,r}(t) = \beta^{r-k} e^{-\alpha t} I_{k-r} + \beta^{r-k} e^{-\alpha t} \sum_{j=0}^{\infty} \{I_{2(j+1)(N+1)+k-r} + I_{2(j+1)(N+1)-k-r}\}$$

$$+ \lambda \beta^{r-k} \sum_{j=0}^{\infty} \int_0^t e^{-\alpha \tau} \{I_{2(j+1)(N+1)-k-r-2} - 2\beta^{-1} I_{2(j+1)(N+1)-k-r-1} + \beta^{-2} I_{2(j+1)(N+1)-k-r}$$

$$+ I_{2j(N+1)+k+r+2} - 2\beta^{-1} I_{2j(N+1)+k+r+1} + \beta^{-2} I_{2j(N+1)+k+r}\} d\tau \quad (r = 0, 1, 2, \dots, N). \quad (4.4)$$

In order to lighten the algebra, we give the expected value for the special case $k = 0$ only:

$$E(r | 0, t) = (\lambda - \mu)t + \lambda \sum_{j=0}^{\infty} \int_0^t \tau^{-1}(t - \tau) e^{-\alpha \tau} \Phi_0(\tau) d\tau, \quad (4.5)$$

$$\text{where } \Phi_0(\tau) = \beta^{-3}\{(2jN - 1)I_{2jN-1} - [2(j+1)N + 1]I_{2(j+1)N+1}\}$$

$$+ \beta^{-2}\{2(j+1)NI_{2(j+1)N} - 2jNI_{2jN}\} + \beta^{-(N-1)}\{[(2j+1)N + 1]I_{(2j+1)N+1}$$

$$- [(2j+1)N - 1]I_{(2j+1)N-1}\}.$$

Asymptotic results are most easily obtained by applying (1.4) to (4.2). Thus the generating function of the stationary distribution is

$$\lim_{s \rightarrow +0} sg_k(z, s) = \lim_{t \rightarrow \infty} G_k(z, t) = \begin{cases} \frac{(\lambda - \mu)\{z^{N+1} - (\mu/\lambda)^{N+1}\}}{(\lambda z - \mu)\{1 - (\mu/\lambda)^{N+1}\}} & \text{if } \lambda \neq \mu, \\ \frac{(1 - z^{N+1})}{(1 - z)(N + 1)} & \text{if } \lambda = \mu, \end{cases} \quad (4.6)$$

and hence

$$\lim_{t \rightarrow \infty} P_{k,r}(t) = \begin{cases} \frac{(\lambda - \mu)(\mu/\lambda)^{N-r}}{\lambda\{1 - (\mu/\lambda)^{N+1}\}} & \text{if } \lambda \neq \mu, \\ 1/(N + 1) & \text{if } \lambda = \mu. \end{cases} \quad (4.7)$$

The expectation of r in the stationary distribution is

$$E(r, \infty) = \begin{cases} \left\{N - \frac{(\mu/\lambda)[1 - (\mu/\lambda)^N]}{1 - \mu/\lambda}\right\} \{1 - (\mu/\lambda)^{N+1}\}^{-1} & \text{if } \lambda \neq \mu, \\ \frac{1}{2}N & \text{if } \lambda = \mu. \end{cases} \quad (4.8)$$

In the context of queueing theory the random walk considered above is the single server queue problem with the restriction that the size of the queue is $\leq N$. In this case new arrivals do not join the queue if there are already $N-1$ 'customers' waiting for service. The probability $P_{k,r}(t)$ given by (4.4) is then the probability of having r 'customers' in the system at time t (including the one being served) conditional on k 'customers' at time zero. We list below the Laplace transform of some formulae of interest. Inversion of these results involves simple but tedious calculations which we do not carry out.

The expected length of the waiting line L at time t , conditional on initial length $k-1$ is

$$E(L | k-1, t) = E(r | k, t) + P_{k,0}(t) - 1$$

and has the Laplace transform

$$\mathcal{L}[E(L | k-1, t)] = (\lambda - \mu)s^{-2} + s^{-1}\{k-1 + Y_k - X_k\} + \frac{\{u_1^k + u_1^N(1-u_1)X_k + u_2^{-1}(u_2-1)Y_k\}}{\lambda(u_2-u_1)}, \quad (4.9)$$

and, by (1.4),

$$E(L | k-1, \infty) = \{N - \{1 - (\mu/\lambda)^N\} (1 - \mu/\lambda)^{-1}\} \{1 - (\mu/\lambda)^{N+1}\}^{-1}.$$

The distribution of the duration of a busy period, $\rho_k(t)$, is obtained by considering the random walk with an absorbing barrier at the origin; i.e. with the boundary conditions

$$g_0 = 1/s, \quad (s + \mu)g_N = \mu g_{N-1} + z^N. \quad (4.10)$$

Proceeding as before the solution of (2.4) with these boundary conditions is

$$g_k = s^{-1}V_k + \phi\{(z^N - z^{N+1})W_k + z^k - V_k\}, \quad (4.11)$$

where

$$V_k = \{u_1^k u_2^N (u_2 - 1) + u_1^N u_2^k (1 - u_1)\} \{u_2^N (u_2 - 1) + u_1^N (1 - u_1)\}^{-1}$$

and

$$W_k = (u_2^k - u_1^k) \{u_2^N (u_2 - 1) + u_1^N (1 - u_1)\}^{-1}.$$

The probability that the particle will be absorbed at the origin before time t , conditional on initial state k , is $\mathcal{L}^{-1}[V_k/s]$. Thus the probability density, $\rho_k(t)$, of the duration of a busy period has the Laplace transform $\mathcal{L}[\rho_k(t)] = V_k(s)$. (4.12)

The formulae are simplified when $N \rightarrow \infty$ and we have the familiar single server queue problem in which no limit is placed on queue length. The appropriate generating function is given by (4.3), and the transition probabilities by

$$P_{k,r}(t) = \beta^{r-k} \left\{ e^{-\alpha t} I_{k-r} + \lambda \int_0^t e^{-\alpha \tau} \{ \beta^{-2} I_{k+r} - 2\beta^{-1} I_{k+r+1} + I_{k+r+2} \} d\tau \right\} \quad (r = 0, 1, 2, \dots). \quad (4.13)$$

(4.13) is identical with the result given by Ledermann & Reuter (1954, p. 366), obtained by the use of spectral theory*. The expected value of r is

$$E(r | k, t) = k + (\lambda - \mu)t + \sum_{j=0}^{\infty} \beta^{-(k+1+j)} (k+1+j) \int_0^t \tau^{-1} e^{-\alpha \tau} I_{k+1+j} d\tau. \quad (4.14)$$

* This result has also been obtained by the following authors: Bailey (1954), by solving the forward equations using a generating function technique; Clarke (1956), by solving an integral equation of Volterra type; Champernowne (1956), by a direct probability argument; and Conolly (1958), by an argument similar to ours. We are indebted to the referee for the last three of these references.

The expected length of the waiting line can be written down from (4.13) and (4.14). The probability density of the duration of a busy period, $\rho_k(t)$, is given by differentiating (3.13), thus

$$\rho_k(t) dt = k\beta^{-k}t^{-1}e^{-zt}I_k dt. \quad (4.15)$$

The ergodic properties of this process have been investigated in detail by Kendall (1951).

In order to obtain the 'diffusion process' limit, set again $Nh = L$ and translate the origin of x and y to $\frac{1}{2}L$. The reflecting boundary condition is seen to correspond to $\partial f / \partial y = 0$ at $\pm \frac{1}{2}L$, and one finds the transform $w(\theta, s, y)$ of $f(x | y, t)$ (using the same notation as in (3.26))

$$w(\theta, s, y) = e^{i\theta y} \psi - \frac{i\theta \psi e^{-ay/\sigma^2}}{2\omega \sinh \omega L} \{e^{(i\theta + a/\sigma^2)\frac{1}{2}L} \cosh \omega(y + \frac{1}{2}L) + e^{-(i\theta + a/\sigma^2)\frac{1}{2}L} \cosh \omega(y - \frac{1}{2}L)\}. \quad (4.16)$$

5. QUEUES WITH N SERVERS

The exponential arrival and service time queue with N servers (Feller, 1957, p. 415) is defined in our notation by the equations

$$(s + \lambda + k\mu)g_k = \lambda g_{k+1} + \mu k g_{k-1} + z^k \quad (k = 0, 1, \dots, N), \quad (5.1)$$

$$(s + \lambda + N\mu)g_k = \lambda g_{k+1} + \mu N g_{k-1} + z^k \quad (k = N, N+1, \dots). \quad (5.2)$$

Before giving the solution $g_k(z, s)$ of the system (5.1) and (5.2) we mention briefly the generalization of (1.3) to the case where the coefficients of the difference equation are linear functions of k ;

$$\{k(a_1 + b_1) + s + a_0 + b_0\}g_k = (ka_1 + a_0)g_{k+1} + (kb_1 + b_0)g_{k-1} + z^k. \quad (5.3)$$

The homogeneous equation obtained from (5.3) is the well-known hypergeometric difference equation, and its solutions have been studied in detail by several authors (e.g. Batchelder, 1927). In fact there are twenty-four such solutions corresponding to the twenty-four solutions of the hypergeometric differential equation (Batchelder, 1927, p. 101), and we may choose two of these appropriate to our problem, say $f_k(s)$ and $h_k(s)$. A particular solution of the non-homogeneous equation may be found by taking the Laplace transform of the generating function $G_k(z, t)$ of the original birth and death process. $G_k(z, t)$ can be found readily by standard methods and has the convoluted binomial form

$$G_k(z, t) = z^{b_0/b_1} \left\{ \frac{b_1(e^{(a_1-b_1)t} - 1) + z(a_1 - b_1 e^{(a_1-b_1)t})}{a_1 - b_1} \right\}^{k-b_0/b_1} \times \left\{ \frac{(a_1 e^{(a_1-b_1)t} - b_1) + za_1(1 - e^{(a_1-b_1)t})}{a_1 - b_1} \right\}^{a_0/a_1 - k}. \quad (5.4)$$

The general solution of (5.3) is then

$$g_k(z, s) = \psi_k(z, s) + A(z, s)f_k(s) + B(z, s)h_k(s), \quad (5.5)$$

where $\psi_k(z, s) = \mathcal{L}[G_k(z, t)]$. As before A and B may be chosen to satisfy given boundary conditions.

There are several interesting variants of (5.3). For example, a linear birth and death process with $a_0 = 0 = b_0$ and reflecting barriers at $k = N_1, N_2$; $N_2 > N_1$, could be used as a linearized model to investigate the logistic process (Kendall, 1949). The essential feature of Kendall's logistic model is that the coefficients are quadratic in k , so that the states N_1 and N_2 are natural reflecting barriers. Explicit solutions can be obtained for the linear boundary value problem and could prove useful as an approximation to the quadratic case.

Since (5.1) and (5.2) are special cases of (5.3) their solutions are, respectively,

$$g_k(z, s) = \psi_k(z, s) + A(z, s) v_k(s) \quad (k = 0, 1, \dots, N), \quad (5.6)$$

$$g_k(z, s) = z^k \phi(z, s) + B(z, s) u_1^k(s) \quad (k = N, N+1, \dots). \quad (5.7)$$

$u_1(s)$ and $\phi(z, s)$ are the same as in previous paragraphs with μN substituted for μ ;

$$\begin{aligned} \psi_k(z, s) &= \mathcal{L}\{[1 - (1-z)e^{-\mu t}]^k \exp[-(\lambda/\mu)(1-z)(1-e^{-\mu t})]\} \\ &= e^{-(\lambda/\mu)(1-z)} \sum_{j=0}^k \binom{k}{j} \frac{(z-1)^j}{\mu} {}_1F_1(s/\mu+j; s/\mu+j+1; (\lambda/\mu)(1-z)). \end{aligned} \quad (5.8)$$

$$\begin{aligned} v_k(s) &= \int_0^1 e^{-\lambda x/\mu} x^k (1-x)^{s/\mu-1} dx \\ &= B(k+1, s/\mu) {}_1F_1(k+1; k+1+s/\mu; -\lambda/\mu), \end{aligned} \quad (5.9)$$

where $B(m, n)$ and ${}_1F_1(m; n; x)$ are the Beta and confluent hypergeometric functions, respectively. The boundary at $k = 0$ being a natural one, A and B are determined by the equation for $g_N(z, s)$, yielding the solution

$$g_k(z, s) = \begin{cases} \psi_k + v_k \frac{\{z^{N-1}(z-u_1)\phi + u_1\psi_{N-1} - \psi_N\}}{v_N - u_1 v_{N-1}}; & (k = 0, 1, \dots, N), \\ z^k \phi + \frac{u_1^{k+1-N} \{z^{N-1}(z v_{N-1} - v_N)\phi + v_N \psi_{N-1} - v_{N-1} \psi_N\}}{v_N - u_1 v_{N-1}}; & (k = N, N+1, \dots) \end{cases} \quad (5.10)$$

The Laplace transform of the expected number of 'customers' in the system at time t , given k initially, is

$$\mathcal{L}[E(r | k, t)] = \begin{cases} \lambda/(\mu s) + \frac{\mu k - \lambda}{\mu(s + \mu)} + R_k \left\{ \frac{(\mu N - \lambda)}{\lambda \mu (u_2 - 1)} + \frac{1}{\lambda (u_2 - 1)^2} - \frac{(1 - u_1)(N - \lambda/\mu) + u_1}{s + \mu} \right\}; & (k = 0, \dots, 1, N), \\ k/s + \frac{(\lambda - \mu)}{s} + S_k \left\{ \frac{(\lambda - \mu)(v_{N-1} - v_N)}{s^2} - \mu \frac{\{(N-1 - \lambda/\mu)v_N - (N - N\mu)v_{N-1}\}}{s(s + \mu)} \right\} & (k = N, N+1, \dots), \end{cases} \quad (5.11)$$

where

$$\begin{aligned} R_k &= v_k [v_N - u_1 v_{N-1}]^{-1}, \\ S_k &= u_1^{k+1-N} [v_N - u_1 v_{N-1}]^{-1}. \end{aligned}$$

The stationary distribution, which exists for $\mu N > \lambda$, may be found directly from (5.10) using (1.4). The calculations involved in taking the limit are considerably simplified by using the following recurrence relation for confluent hypergeometric functions (Sneddon, 1956)

$$(x/\gamma) {}_1F_1(\alpha+1; \gamma+1; x) = {}_1F_1(\alpha+1; \gamma; x) - {}_1F_1(\alpha; \gamma; x). \quad (5.12)$$

Using (5.12) repeatedly in the terms $u_1 \psi_{N-1} - \psi_N$ and $v_N - u_1 v_{N-1}$, and expanding as a power series, we have

$$\begin{aligned} \lim_{s \rightarrow +0} s g_k(z, s) &= \lim_{t \rightarrow \infty} G_k(z, t) \\ &= P_0 \left\{ \sum_{r=0}^{N-1} \frac{(\lambda/\mu)^r}{r!} z^r + \sum_{r=N}^{\infty} \frac{(\lambda/\mu)^r}{N! N^{r-N}} z^r \right\}; \quad \mu N > \lambda, \end{aligned} \quad (5.13)$$

where

$$P_0 = \left\{ \sum_{r=0}^{N-1} \frac{(\lambda/\mu)^r}{r!} + \sum_{r=N}^{\infty} \frac{(\lambda/\mu)^r}{N! N^{r-N}} \right\}^{-1}.$$

(5.13) agrees with the result given by Feller (1957, p. 415).

The barrier at the origin of the system (5.1) and (5.2) is a natural reflecting one. By imposing an absorbing barrier at the origin

$$g_0 = 1/s \quad (5.14)$$

and solving (5.1), (5.2), for $k > 0$, and (5.14) for $k = 0$, we can find the probability density $\rho_k(t)$, of the first passage time through the origin, i.e. the density of the duration of a busy period. Say this solution is $g_k = g_k(z, s)$, $k \geq 1$. Then

$$\rho_k(t) = \frac{dP_{k0}(t)}{dt} = \mathcal{L}^{-1}\{sg_k(0, s)\}. \quad (5.15)$$

[Note that $\int_0^\infty \rho_k(t) dt = 1$ if and only if $\mu N > \lambda$.] The moments of the distribution $\rho_k(t)$ may be found directly from (5.15) using the following elementary properties of the Laplace transform: If $\mathcal{L}\{f(t)\} = g(s)$ then

$$\left. \begin{aligned} \mathcal{L}\{t^n f(t)\} &= (-1)^n \frac{d^n g(s)}{ds^n}, \\ \mathcal{L}\left\{\int_0^t f(\tau) d\tau\right\} &= s^{-1}g(s). \end{aligned} \right\} \quad (5.16)$$

and

Using (5.16) and (1.4) the expected value of t^n conditional on k , $E(t^n | k)$, is

$$E(t^n | k) = \lim_{t \rightarrow \infty} \int_0^t \tau^n \rho_k(\tau) d\tau = \left\{ (-1)^n \frac{d^n}{ds^n} (sg_k(0, s)) \right\}_{s=0}. \quad (5.17)$$

The solution of this absorbing barrier problem follows the same lines as before, with the modification that we require the general solution of (5.1), $k > 0$, with two arbitrary constants. It is easily verified that a second solution of the homogeneous equation obtained from (5.1) is

$$w_k(s) = \sum_{j=0}^k \binom{k}{j} (\mu/\lambda)^j \Gamma(j + s/\mu). \quad (5.18)$$

The general solution of (5.1), $k = 1, 2, \dots, N$, is then

$$g_k(z, s) = \psi_k(z, s) + A(z, s) v_k(s) + C(z, s) w_k(s), \quad (5.19)$$

where ψ_k and v_k are given by (5.8) and (5.9) respectively. C can be eliminated by substituting (5.19) in (5.14) to give

$$g_k = \{w_k + s(w_0 \psi_k - w_k \psi_0)\} (sw_0)^{-1} + A(w_0 v_k - w_k v_0) w_0^{-1}. \quad (5.20)$$

Elimination of the constants A and B from (5.20) and (5.6) gives the solution $g_k(z, s)$. We give the expression for the required quantity $g_k(0, s)$ for $k \geq N$ only:

$$g_k(0, s) = \frac{u_1^{k+1-N} \{(w_0 v_N - w_N v_0) (w_{N-1} + s w_0 \psi_{N-1} - s w_{N-1} \psi_0) - (w_0 v_{N-1} - w_{N-1} v_0) (w_N + s w_0 \psi_N - s w_N \psi_0)\}}{s w_0 \{(w_0 v_N - w_N v_0) - u_1 (w_0 v_{N-1} - w_{N-1} v_0)\}}, \quad (5.21)$$

where the function $\psi(z, s)$ appearing in (5.21) has been evaluated at $z = 0$.

Application of (5.17) to (5.21) yields the moments of the duration of a busy period for $k \geq N$. Results for $1 \leq k < N$ may of course be obtained by the same method.

REFERENCES

- BAILEY, N. T. J. (1954). A continuous time treatment of a simple queue using generating functions. *J. R. Statist. Soc. B*, **16**, 288-91.
- BARTLETT, M. S. (1956). *An Introduction to Stochastic Processes*. Cambridge University Press.
- BATCHELDER, P. M. (1927). *An Introduction to Linear Difference Equations*. Harvard University Press.
- CHAMPERNOWNE, D. G. (1956). An elementary method of solution of the queueing problem with a single server and constant parameters. *J. R. Statist. Soc. B*, **18**, 125-8.
- CHANDRASEKHAR, S. (1943). Stochastic processes in physics and astronomy. *Rev. Mod. Phys.* **15**, 1-89.
- CLARKE, A. B. (1956). A waiting line process of Markov type. *Ann. Math. Statist.* **27**, 452-9.
- CONOLLY, B. W. (1958). A difference equation technique applied to the simple queue. *J. R. Statist. Soc. B*, **20**, 165-7.
- ERDELYI, A. (1954). *Tables of Integral Transforms*, **1**. New York: McGraw-Hill.
- FELLER, W. (1957). *An Introduction to Probability Theory and its Applications*, 2nd ed. New York: John Wiley and Sons.
- IRWIN, J. O. (1937). The frequency distribution of the difference between two independent variates following the same Poisson distribution. *J. R. Statist. Soc.* **100**, 415.
- KARLIN, S. & MCGREGOR, J. (1958). Many server queueing processes with Poisson input and exponential service times. *Pacific J. Math.* **8**, 87-118.
- KENDALL, D. G. (1949). Stochastic processes and population growth. *J. R. Statist. Soc. B*, **11**, 230-64.
- KENDALL, D. G. (1951). Some problems in the theory of queues. *J. R. Statist. Soc. B*, **13**, 151-85.
- LEDERMANN, W. & REUTER, G. E. H. (1954). Spectral theory for the differential equations of simple birth and death processes. *Phil. Trans. A*, **246**, 312-69.
- MORSE, P. M. (1958). *Queues, Inventories and Maintenance*. New York: John Wiley and Sons.
- MOYAL, J. E. (1957). Discontinuous Markov processes. *Acta Mathematica*, **98**, 221-64.
- MOYAL, J. E. (1959). Incomplete discontinuous Markov processes. Submitted to *Quart. J. Math.* Oxford.
- SNEDDON, I. N. (1956). *Special Functions of Mathematical Physics and Chemistry*. Edinburgh: Oliver and Boyd Ltd.
- WATSON, G. N. (1944). *Theory of Bessel Functions*, 2nd ed. Cambridge University Press.
- WIDDER, D. V. (1946). *The Laplace Transform*. Princeton University Press.

NOTE ON THE COMPARISON OF SEVERAL REALIZATIONS OF A MARKOFF CHAIN

By J. H. DARWIN

Department of Scientific and Industrial Research, Wellington, New Zealand

1. INTRODUCTION AND SUMMARY

Consider a probability chain defined by $X(t)$ where t takes in succession the values $t_1, t_2, \dots$, where $t_{i+1} - t_i = \tau, \tau > 0$. Suppose the chain can be in any of s states at any time t_i . We describe the event 'the chain is in state j when $t = t_i$ ' by saying $X(t_i) = j$. Then the chain is called u -dependent if $\text{Prob}(X(t_i) = j)$ is a function of the u values

$$X(t_i - \tau), \quad X(t_i - 2\tau), \quad \dots, \quad X(t_i - u\tau).$$

We suppose this function is independent of i .

A great deal of research into u -dependent chains of this kind has been carried out in recent years. Bartlett (1951) discussed the asymptotic distribution theory for functions of long realizations from such chains, that is for realizations consisting of $X(t_1), \dots, X(t_N)$ when N is large. Following this preliminary work, writers such as Hoel (1954), Whittle (1955), Good (1955), Anderson & Goodman (1957) and Goodman (1958) have developed asymptotic tests of significance which, for instance, compare the maximum likelihood ($m.l$) estimates of the probability $\{X(t_i + r\tau) = j \mid \text{any set } X(t_i + (r-1)\tau), \dots, X(t_i)\}$ with a given set of values of these probabilities, which examine the length u of dependence of the chain, which are suitable when s is large, etc.

In this note we discuss a one-dependent or Markoff chain with s states. We suppose $\{X(t_i) = k \mid X(t_i - \tau) = j\} = p_{jk}$. We shall refer to p_{jk} as a transition probability. We suppose also there is a unique occupation probability P_j by which is meant $P\{X(t_N) = j\}$ tends to P_j as N tends to infinity. We suppose P_j and p_{jk} exist for all j and k .

We develop:

(a) A test of the equality of r sets of transition probabilities p_{jk} . This test follows automatically from Bartlett's work.

(b) A test of the equality of r sets of transition probabilities p_{jj} separated from a test of the equality of r sets $p_{jk}/(1 - p_{jj}), j \neq k$. This division of the test (a) is useful when τ is small, in that it then throws light on the reasons for the differences in occupation times of the s states from set to set.

(c) A test of the equality of r sets of occupation probabilities P_j . This is of interest when τ is not small. The test criterion is difficult to calculate on a desk machine, but not on most modern electronic computers. We show, when $s = 2$, the direction of inadequacy of the assumption of independence of $X(t_i)$ and $X(t_i + \tau)$ when $X(t_i)$ is a Markoff chain, and a comparison is being made of the occupation numbers for two long realizations.

2. THE COMPARISON OF r SETS OF TRANSITION PROBABILITIES

Bartlett (1951) showed that:

(i) n_{jk} the number of values of i for which $X(t_i) = j, X(t_i + \tau) = k$ or, say, the number of transitions from state j to state k is for a long realization approximately normally dis-

tributed with mean $NP_j p_{jk}$ and variance $NP_j p_{jk}(1-p_{jk})$, where N is the length of the realization; $\text{cov}(n_{jk}, n_{jl})$ is $-NP_j p_{jk} p_{jl}$. We suppose all P_j are greater than 0;

(ii) if $n_{j.} = \sum_k n_{jk}$, $n_{j.}$ is for large N distributed approximately normally with mean NP_j and variance proportional to N , so that $n_{j.}/(NP_j)$ tends to one in probability;

(iii) the maximum likelihood estimate of $p_{jk}, n_{jk}/n_{j.}$ is for large N uncorrelated with n_{lm}/n_l , if $j \neq l$.

It follows from (i) that for large N the s values n_{jk} , $k = 1, \dots, s$ have a multinomial distribution with index NP_j and parameters p_{jk} , $k = 1, \dots, s$; from (iii) that these s multinomials are independent of each other; and from (ii) and a theorem of Cramér's (1946, p. 254) on the replacement of parameters by quantities which tend to them in probability, that in the multinomial for the s values n_{jk} , $k = 1, \dots, s$ we may replace NP_j by $n_{j.}$.

We consider testing the hypothesis that r sets of values p_{jk} are the same, that is that p_{jkl} where $l = 1, \dots, r$, is independent of l for every combination j, k . The data are r long realizations one from each of the respective chains. We are on very familiar ground because the preceding paragraph implies that the test must be related to a combination of tests of $s \times r \times s$ contingency tables, each table corresponding to a value j . The likelihood ratio criterion (l.r.c.) to test the hypothesis is

$$2[\sum n_{jkl} \log_e n_{jkl} - \sum n_{j.l} \log_e n_{j.l} - \sum n_{jk.} \log_e n_{jk.} + \sum n_{j..} \log_e n_{j..}]. \quad (1)$$

In this and later formulae a dot in place of a suffix means summation has been carried out over the replaced variable. Thus, for example, $n_{j..} = \sum_{k,l} n_{jkl}$. The remaining summation signs indicate that summation is to be carried out over all possible values of the remaining suffix variables: $n_{j.} \sim n_{.j}$, with equality for all j , if the last state is the same as the first.

It is easily proved by a standard expansion of $(n_{jkl} - n_{j.l} p_{jkl})/\sqrt{n_{j.l}}$ that to first order in the $n_{j.l}$ this criterion has the same value as the sum of $s \chi^2$ -type criteria for $s \times r \times s$ contingency tables. The distribution of the criterion (1) therefore is that of χ^2 with $s(r-1)(s-1)$ degrees of freedom.

3. THE COMPARISON OF TRANSITION PROBABILITIES FOR SMALL τ

Suppose the chain is the result of observation of a process defined in continuous time. We now take the parameter t of § 1 to refer to time. If τ is decreased it may happen that the length of dependence of the resulting chain is increased or that any type of u -dependence disappears. We consider only chains which are still Markoff chains for such low values of τ that it is reasonable to suppose no change of state has been missed—that is that there has been at most one change of state between times t_i and $t_i + \tau$.

That such a case may occur in practice is shown by the example in educational theory that prompted the work of this paper (see § 5 below). An educator defines a number of states in which a class may be: teacher giving formal instruction, pupil asking a question, etc. This leads to a chain, approximately Markovian, even for low τ . If τ is smaller than the average amount of time spent in the state in which the chain stays for the shortest time it is unlikely that any change of state will have been missed.

It then becomes possible to make a simple interpretation of comparisons between sets of values p_{jj} . We suppose now

$$\{X(t_i + \tau) = j = X(t_i + 2\tau) = \dots = X(t_i + (n-1)\tau); X(t_i + n\tau) = k_2 \mid X(t_i) = k_1\} \\ = (1 - p_{k_1 j}) v_{k_1 j} p_{jj}^{n-1} (1 - p_{jj}) v_{jk_2}. \quad (2)$$

Then $(1 - p_{jj})v_{jk}$ plays the role of p_{jk} , $j \neq k$, in the Markoff chain originally considered, and the use of v_{jk} amounts to a change of parameters for that chain. The quantity v_{jk} is a conditional probability, the probability that if the chain has moved from state j it has moved to state k . The probability that the chain stays in state j for $n - 1$ further observations once it has reached j is $(1 - p_{jj})p_{jj}^{n-1}$. This geometric distribution can then be described as the distribution of time spent in state j . It corresponds to an exponential distribution $\exp(-\lambda_j t) \lambda_j dt$ of time spent in state j for the corresponding process defined in continuous time.

It is commonly desired to compare the occupation times (by the occupation time of state j we mean the number of times t the state j was observed) for two or more long realizations. For chains of independent observations this comparison is made by an ordinary contingency table analysis. For values of τ not small enough to enable us to make the special assumptions of this section, the nearest equivalent is a test comparing values of occupation probabilities P_j . We discuss such a test in § 4. When the assumptions of this section hold, the test of § 4 and the test of § 2 can be supplemented by a separate test of the equality of the sets p_{jj} . Such a test will be useful for instance when two sets of occupation times are about equal, but for one realization there has been a rapid fluctuation from state to state, while for the other there has been a slower rate of change with longer periods spent in each state. The likelihood ratio criterion to test the hypothesis that r sets v_{jkl} , $l = 1, \dots, r$, are equal regardless of the values of the sets p_{jil} is

$$2\left[\sum_{j \neq k} n_{jkl} \log_e n_{jkl} - \sum (n_{j..l} - n_{jjl}) \log_e (n_{j..l} - n_{jjl}) - \sum_{j \neq k} n_{jk.} \log_e n_{jk.} + \sum (n_{j..} - n_{jj.}) \log_e (n_{j..} - n_{jj.})\right]. \quad (3)$$

This is of the same form as (1) with $n_{j..l} - n_{jjl}$ replacing $n_{j..l}$. In considering (1) we were able to suppose the n_{jkl} , $k = 1, \dots, s$ had a multinomial distribution with index $n_{j..l}$ for large N_l . If now we consider n_{jjl} fixed in this multinomial distribution the conditional distribution of the n_{jkl} , $j \neq k$, for large N_l is again multinomial with index $n_{j..l} - n_{jjl}$ and parameters $p_{jkl}/(1 - p_{jjl})$. It follows as in § 2 by analogy with the analysis of an $r \times (s - 1)$ contingency table that the distribution of (3) for large $n_{j..l} - n_{jjl}$ is χ^2 with $s(r - 1)(s - 2)$ degrees of freedom. Since this is independent of the $n_{j..l} - n_{jjl}$ provided they are large, it is the general asymptotic distribution of (3). These degrees of freedom virtually follow from the fact that the $s(s - 1)$ values $p_{jkl}/(1 - p_{jjl})$, $j \neq k$, have one linear restriction for each value of j .

The likelihood ratio criterion to test the hypothesis that r sets p_{jil} , $l = 1, \dots, r$, are equal regardless of the value of the sets v_{jkl} is

$$2\left[\sum (n_{j..l} - n_{jjl}) \log_e (n_{j..l} - n_{jjl}) + \sum n_{jjl} \log_e n_{jjl} - \sum n_{j..l} \log_e n_{j..l} - \sum (n_{j..} - n_{jj.}) \log_e (n_{j..} - n_{jj.}) - \sum n_{jj.} \log_e n_{jj.} + \sum n_{j..} \log_e n_{j..}\right]. \quad (4)$$

For large $n_{j..l}$ this latter criterion is equivalent to the sum of χ^2 criteria from $s \times 2$ tables, since for given $n_{j..l}$, n_{jjl} is a binomial variable with index $n_{j..l}$ and parameter p_{jjl} , this distribution being independent of the distributions of n_{kkl} , $k \neq j$ for large N_{jl} , N_{kl} . The two criteria (3) and (4) add to give the criterion (1). In the author's view the exact additivity of likelihood ratio criteria over complementary hypotheses of the main hypothesis that r sets p_{jkl} are equal, is an excellent argument for using them rather than the χ^2 type criteria to which they are equivalent for long realizations. These latter criteria annoyingly do not accurately possess this additivity.

For completeness we give the test of the equality of sets λ_{jl} , $l = 1, \dots, r$ of the continuous process when T_{jl} is the amount of time the l th realization passes in state j . The likelihood ratio criterion is

$$2[\sum(n_{j,l} - n_{jil}) \log_e(n_{j,l} - n_{jil}) - \sum(n_{j..} - n_{jil}) \log_e(n_{j..} - n_{jil}) - \sum(n_{j,l} - n_{jil}) \log_e T_{jl} + \sum(n_{j..} - n_{jil}) \log_e T_{j.}]. \quad (5)$$

This is, for large $n_{j,l} - n_{jil}$ distributed as χ^2 with $s(r-1)$ degrees of freedom.

4. THE COMPARISON OF r SETS OF OCCUPATION PROBABILITIES

If for general Markoff chains a test of the equality of r sets p_{jkl} , $l = 1, \dots, r$, had shown no over-all significant differences between the sets it might be considered illogical to test differences between occupation probabilities P_{jl} since these can be expressed in terms of the p_{jkl} . It is, however, conceivable that the sets p_{jkl} may differ on the whole so slightly that the criterion (1) is found to be not significant, but that the particular functions of them represented by the P_{jl} may show important differences from one chain to another. Accordingly, we develop an asymptotic test of the differences between two sets of occupation probabilities P_{jl} , starting from the fact (Bartlett, 1951) that the $n_{j.}$ are normally distributed with means NP_j for one realization. We then give the obvious extension to r sets P_{jl} .

We begin by finding the covariance matrix of the $n_{j.}$ for a single long realization of length N from a chain with transition probabilities p_{jk} . Patankar (1954) has shown that

$$\left. \begin{aligned} \text{var } n_{j.} &\sim \sum_{v=1}^{N-1} 2(N-v) P_{jj}^{(v)} + NP_j - N^2 P_j^2, \\ \text{cov}(n_{j.}, n_{k.}) &\sim \sum_{v=1}^{N-1} (N-v) P_{jk}^{(v)} + \sum_{v=1}^{N-1} (N-v) P_{kj}^{(v)} - N^2 P_j P_k, \end{aligned} \right\} \quad (6)$$

where $P_{jk}^{(v)}$ is the stable probability of the chain being in state j when $t = t_i$ and in state k when $t = t_i + v\tau$. Suppose the matrix of transition probabilities p_{jk} is called T . We shall consider that the roots of T are distinct. We are already assuming that the only root of modulus one is one. Then T has a spectral resolution (see, for example, Bartlett, 1955, p. 26)

$$T = X + \sum_{\lambda \neq 1} \lambda P(\lambda), \quad (7)$$

where X is

$$\begin{pmatrix} P_1 & P_2 & \dots & P_s \\ \vdots & \vdots & & \vdots \\ P_1 & P_2 & \dots & P_s \end{pmatrix},$$

λ is a root of T and $P(\lambda)$ is a matrix $p_{jk}(\lambda)$ of order $s \times s$. The quantity $\sum_v P_{jk}^{(v)}(N-v)$ is then expressible as

$$P_j(\frac{1}{2}N(N-1)P_k + \sum_{v=1}^{N-1} \sum_{\lambda \neq 1} (N-v) \lambda^v p_{jk}(\lambda)). \quad (8)$$

Since λ is less than one (Fréchet, 1938, p. 105) this is asymptotic for large N to

$$\frac{1}{2}N(N-1)P_j P_k + NP_j \sum_{\lambda \neq 1} p_{jk}(\lambda) \lambda / (1-\lambda). \quad (9)$$

Introducing this into (6), we may write

$$\begin{aligned} \text{var } n_{j.} &\sim N(2P_j \times (j, j) \text{th element of } (I - T + X)^{-1} - P_j - P_j^2) \\ \text{cov}(n_{j.}, n_{k.}) &\sim N(P_j \times (j, k) \text{th element of } (I - T + X)^{-1} \\ &\quad + P_k \times (k, j) \text{th element of } (I - T + X)^{-1} - P_j P_k). \end{aligned} \quad (10)$$

We have replaced $\sum_{\lambda=1} p_{jk}(\lambda)/(1-\lambda)$ by $(I - T + X)^{-1}$; this replacement is permissible since in this sum λ is less than one. If we write D for the diagonal matrix, $\text{diag}(P_1, \dots, P_s)$, we may write for the covariance matrix of the n_j ,

$$N[D(I - T + X)^{-1} + (D(I - T + X)^{-1})' - D(I + X)]. \quad (11)$$

The matrix (11) is singular since $\sum n_j = N$. We suppose the first $s-1$ of the n_j have a non-singular distribution which we write as

$$\frac{1}{((2\pi N)^{s-1} |A|)^{\frac{1}{2}}} \exp \left[-\frac{1}{2} \left(\frac{n_{j.}}{N} - P_j \right)' N A \left(\frac{n_{j.}}{N} - P_j \right) \right] dn_1, \dots, dn_{s-1}, \quad (12)$$

where the row vector $n_1./N - P_1, \dots, n_{s-1.}/N - P_{s-1}$ is written $(n_{j.}/N - P_j)'$. Then, under the assumption of a common known A for the two chains being compared, it easily follows that the likelihood ratio interior to test the hypothesis that the set P_j for the realization with occupation numbers $n_1, \dots, n_{s-1.}$ is equal to the set P_j for the realization with occupation numbers $m_1, \dots, m_{s-1.}$ is

$$\frac{MN}{M+N} \left(\frac{m_{j.}}{M} - \frac{n_{j.}}{N} \right)' A \left(\frac{m_{j.}}{M} - \frac{n_{j.}}{N} \right). \quad (13)$$

The similar criterion for testing a single set of occupation probabilities against a hypothetical set was suggested by P. S. Neal (see Patankar, 1954). The joint frequency function of the m_j and the n_j may be expressed as

$$\text{const. exp} \left[\frac{-MN}{2(M+N)} \left(\frac{m_{j.}}{M} - \frac{n_{j.}}{N} \right)' A \left(\frac{m_{j.}}{M} - \frac{n_{j.}}{N} \right) - \frac{1}{2(M+N)} \left(N \left(\frac{n_{j.}}{N} - P_j \right) + M \left(\frac{m_{j.}}{M} - P_j \right) \right)' A \left(N \left(\frac{n_{j.}}{N} - P_j \right) + M \left(\frac{m_{j.}}{M} - P_j \right) \right) \right]. \quad (14)$$

The non-singular transformation

$$\left. \begin{aligned} z_j &= n_{j.}/N - P_j - (m_{j.}/M - P_j), \\ w_j &= N(n_{j.}/N - P_j) + M(m_{j.}/M - P_j), \end{aligned} \right\} \quad (15)$$

enables the criterion (13) to be expressed purely in terms of z_j . The exponent in (14) is in fact the sum of two quadratics, one in z_j of rank $s-1$ equal to minus half the criterion (13), the other in w_j also of rank $s-1$. It follows that the criterion (13) is asymptotically distributed as χ^2 with $s-1$ degrees of freedom when A is known.

However, since A is not known it has to be replaced by an estimate. A is a function of the p_{jk} . We replace p_{jk} by the pooled likelihood estimate $(m_{jk} + n_{jk})/(m_{j.} + n_{j.})$. Under the null hypothesis of equal sets p_{jk} this estimate is for large $N + M$ distributed normally with mean p_{jk} and variance proportional to $1/(N + M)$. It follows that the estimate tends in probability to p_{jk} . Since A is a finite function of the s^2 values p_{jk} , $A(p_{jk})$ also tends in probability to its true value as $M + N$ becomes large. By a theorem of Cramér (1946, p. 254) this replacement of A by such an estimate will not affect the limiting distribution of the criterion (13).

Until recently it would have been of purely academic interest to suggest testing a criterion the calculation of which even by a computer using a desk machine is extremely laborious. But with the growth in numbers of electronic computers, most of which have inversion and multiplication of matrices in their libraries, calculation of the criterion (13) will often be feasible and cheap (the impetus to provide a test of occupation was in fact provided by an organization with access to a high speed computer).

Examination of the criterion (13) for a small number s of states throws light on the behaviour of the χ^2 test of a contingency table of occupation numbers when the chain is one-dependent and not zero-dependent as required for the validity of such a test. We discuss the simplest example $s = 2$.

4.1. Two states; $s = 2$

T is now $\begin{pmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \end{pmatrix}$ and X is $\begin{pmatrix} P_1 & P_2 \\ P_1 & P_2 \end{pmatrix}$,

where $P_1 = p_{2,1}/(p_{1,2} + p_{2,1})$ and $P_2 = p_{1,2}/(p_{1,2} + p_{2,1})$.

$|I - T + X| = p_{1,2} + p_{2,1}$ and if the matrix (11) is called VN

$$V = \frac{P_1 P_2 (p_{1,1} + p_{2,2})}{p_{1,2} + p_{2,1}} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

It follows that

$$\frac{MN(n_{j.}/N - m_{j.}/M)^2 (p_{1,2} + p_{2,1})}{(M + N) P_1 P_2 (p_{1,1} + p_{2,2})} \quad (16)$$

is asymptotically distributed as χ^2 with 1 degree of freedom.

In calculating this criterion we use the maximum likelihood estimates

$$p_{1,2} = (m_{1,2} + n_{1,2})/(m_{1.} + n_{1.}), \quad p_{2,1} = (m_{2,1} + n_{2,1})/(m_{2.} + n_{2.}),$$

$$p_{1,1} = (m_{1,1} + n_{1,1})/(m_{1.} + n_{1.}) \quad \text{and} \quad p_{2,2} = (m_{2,2} + n_{2,2})/(m_{2.} + n_{2.}).$$

Since $m_{1,2} \sim m_{2,1}$ and $n_{1,2} \sim n_{2,1}$ we may substitute

$$(m_{1.} + n_{1.})/(M + N) \quad \text{and} \quad (m_{2.} + n_{2.})/(M + N)$$

for P_1 and P_2 respectively, where $M = m_{1.} + m_{2.}$, and $N = n_{1.} + n_{2.}$. We may compare this criterion (16) with the usual criterion for the 2×2 table

$m_{1.}$	$m_{2.}$
$n_{1.}$	$n_{2.}$

$$\frac{(M + N) (n_{1.} m_{2.} - n_{2.} m_{1.})^2}{MN(m_{1.} + n_{1.})(m_{2.} + n_{2.})} \quad (17)$$

This latter criterion is

If P_1 and P_2 in (16) are replaced by their maximum likelihood estimates, the only difference between the criteria is the extra factor $(p_{1,2} + p_{2,1})/(p_{1,1} + p_{2,2})$ in (16). Thus if the chain were one-dependent but the χ^2 criterion for a 2×2 table were used as though there were zero-dependence, the expected value of the criterion would not be one but $(p_{1,1} + p_{2,2})/(p_{1,2} + p_{2,1})$. This factor may seriously disturb the validity of the deductions drawn. For instance in the educational example referred to in § 3, the educator may define only two states which may, for example, be teacher-initiated and pupil-initiated activity. An interval of time τ between observations smaller than the average time spent in these states will probably mean that p_{11} and p_{22} dominate p_{12} and p_{21} respectively. Then in using χ^2 one would run the risk of claiming more significance than is really there.

4.2. The comparison of r sets of occupation probabilities.

Suppose it is desired to test the hypothesis that r long realizations are of Markoff chains with the same set of occupation probabilities. Then by a simple extension of the argument of § 4 we can show that the likelihood ratio criterion is

$$\sum_i N_i x_i' A x_i - \left(\sum_i N_i \right) \bar{x}' A \bar{x} = \sum_i N_i (x_i - \bar{x}) A (x_i - \bar{x}). \quad (18)$$

In this the summations are taken over the r realizations for each of which there is an $s \times 1$ vector x_l of quantities of the type $n_{j.}/N$ or $m_{j.}/M$ of § 4.1; N_l is the length of the l th realization and $\bar{x} = \sum N_l x_l / \sum N_l$. The form of A is given by (11)/ N with the maximum likelihood estimate $n_{jk.}/n_{j.}$ substituted for p_{jk} . The criterion (18) is distributed as χ^2 with $(r-1)(s-1)$ degrees of freedom.

5. EXAMPLE

The only example on which the methods of this paper have been tried is provided by data arising from the researches of Dr Ned A. Flanders of the University of Minnesota. These data exhibit such extreme differences that a statistical analysis is scarcely needed. Nevertheless, calculation of the criteria of the previous paragraphs reinforces the argument of § 4.1 and also shows the usefulness of the tests described in § 3 when they are applicable.

Dr Flanders describes six different states in which a class may be, as follows:

- (1) Teacher accepts ideas, feelings and gives praise so as to integrate student behaviour in the problem solving process.
- (2) Teacher asks question.
- (3) Teacher lectures or gives his own opinions and ideas.
- (4) Teacher gives directions, criticism or justifies the use of his authority.
- (5) Student talks.
- (6) None of the above; including silence, confusion and administrative-routine statements made by the teacher.

It is thought that the quality of the teacher is shown by the way in which his classroom switches from state to state as well as by the proportion of time the class spends in each state. Differences between teachers for particular transition probabilities may of course be tested. The comparison of *complete* sets of transition probabilities and occupation probabilities provides a numerical summary of teacher differences which, it is hoped, will be useful in providing a check on the usual subjective summary.

The states in which two classrooms were in, were observed at short regular intervals. Tests of the length of dependence in a chain of such observations suggest that it is approximately true to suppose the observations come from a Markoff chain so that the methods of the paper are applicable. The numbers of observations in the different states are large enough to inspire confidence in the use of our asymptotic tests.

The data are, for the two classrooms:

State ...	1	2	3	4	5	6	1	2	3	4	5	6
1	22	11	24	2	2	7	3	6	9	3	0	8
2	5	23	15	3	42	6	1	9	3	12	27	5
3	4	21	190	25	20	34	2	9	208	32	5	18
4	0	2	14	56	14	28	0	14	32	108	40	40
5	32	15	20	10	56	14	22	14	9	26	224	14
6	5	22	31	18	13	134	1	5	13	53	13	116

The criterion (1) to test the hypothesis that the two sets of transition probabilities are the same has the value 161.0 which on 30 degrees of freedom is very significant. This suggests that the test of the equality of occupation probabilities will be difficult to interpret. We

shall, nevertheless, calculate the required criterion in order to compare it with the criterion calculated on the assumption that successive observations in the classroom are statistically independent.

The estimate $\hat{P}'$ of the row vector of occupation probabilities is the row vector of quantities $(m_j + n_j)/(M + N)$. This is (0.04746, 0.07387, 0.27789, 0.17025, 0.22309, 0.20743). This estimate satisfies the equation $\hat{P}'\hat{T} = \hat{P}'I$ to within one figure in the fifth decimal place, where $\hat{T}$ is equal to T with $(m_{jk} + n_{jk})/(m_j + n_j)$ for p_{jk} . The estimate of the covariance matrix of occupation numbers for a realization of length N is, as in (11),

$$N \times \begin{pmatrix} 0.06711 & 0.01035 & -0.03182 & -0.03164 & 0.01934 & -0.03334 \\ & 0.09545 & -0.06514 & -0.03412 & 0.03807 & -0.04460 \\ & & 0.76613 & -0.14075 & -0.31241 & -0.21599 \\ & & & 0.30988 & -0.10109 & -0.00227 \\ & & & & 0.55254 & -0.19644 \\ & & & & & 0.49266 \end{pmatrix}.$$

The matrix A required by (12) may be found by inverting, for example, the bottom right-hand 5×5 matrix in the above 6×6 matrix. The criterion (13) for testing the equality of the sets of occupation probabilities is then 60.1. The χ^2 criterion for the 2×6 contingency table of values m_j and n_j is 126.0. Thus this example shows the great reduction that can occur in the criterion to test this hypothesis when independence of successive observations is not assumed and when p_{jj} dominates p_{jk} , j not equal to k . This situation for $s = 2$ was discussed in § 4.1.

The assumptions of § 3 are known to be approximately true for data of this kind. Hence we may obtain further information on the factors affecting the great difference between the sets of occupation probabilities by considering the tests described in that paragraph. The criterion for testing the equality of the sets of conditional probabilities v_{jk} , j not equal to k , has the value 95.0 on 24 degrees of freedom, and the criterion for testing the equality of the sets p_{jj} has the value 66.0 on 6 degrees of freedom. Thus it appears that differences in the occupation probabilities are strongly affected by differences in the probabilities of change from one state to another, and more strongly (per degree of freedom involved) by the differences in the length of time spent in a state before a change occurs.

REFERENCES

- ANDERSON, T. W. & GOODMAN, LEO A. (1957). Statistical inference about Markov chains. *Ann. Math. Statist.* **18**, 89–110.
- BARTLETT, M. S. (1951). The frequency goodness of fit for probability chains. *Proc. Camb. Phil. Soc.* **47**, 86–95.
- BARTLETT, M. S. (1955). *An Introduction to Stochastic Processes*. Cambridge University Press.
- CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.
- FRÉCHET, M. (1938). *Recherches théoriques modernes sur la théorie des probabilités* (ed. Borel), **1**, no. 3. Paris.
- GOOD, I. J. (1955). The likelihood ratio test for Markoff chains. *Biometrika*, **42**, 531–33.
- GOODMAN, LEO A. (1958). Simplified run tests and likelihood ratio tests for Markoff chains. *Biometrika*, **44**, 181–97.
- HOEL, P. G. (1954). A test for Markoff chains. *Biometrika*, **44**, 168–77.
- PATANKAR, V. N. (1954). The goodness of fit of frequency distributions obtained from stochastic processes. *Biometrika*, **41**, 450–62.
- WHITTLE, P. (1955). Some distribution and moment formulae for the Markoff chain. *J. R. Statist. Soc. B*, **17**, 235–42.

OVERFLOW AT A TRAFFIC LIGHT

FRANK A. HAIGHT

*Institute of Transportation and Traffic Engineering,
University of California, Los Angeles*

1. LIGHTS IN GENERAL

Suppose vehicles of uniform length arrive at an intersection controlled by a traffic light, with arrival times which constitute a homogeneous Poisson process, with parameter λ . The effect of acceleration will be ignored; once a vehicle is near the traffic light it travels with a uniform speed S unless stopped. When a queue of vehicles is stopped by the light, headway (distance separation between corresponding parts of adjacent vehicles) is constant, and when the cars move off the headway will be a larger constant, denoted by K .

The letter β will be used to indicate the length of the red phase, so that the expected number of arrivals during a red phase will be $\lambda\beta$ for a fixed cycle light and $\lambda E(\beta)$ for a variable cycle light. During the green phase, when there is a queue of cars waiting, they are discharged with time separation $K/S = T$. Once the queue present at the beginning of the green phase has been emptied, the Poisson arrivals continue through the intersection without delay for the remainder of the green phase. Instead of characterizing the green phase by its length, we will use instead an integer N , which is the largest multiple of T contained in its length:

$$\text{length green} = \alpha = NT + \theta T \quad (0 < \theta < 1).$$

Thus, with a sufficiently long queue, at most N vehicles can be discharged during a green phase. N is, however, by no means the maximum number of vehicles which may pass through the intersection during a green phase, for once the queue is dissipated, the Poisson stream may send cars through with time separation $< T$. In some circumstances N will be a constant, and in others a stochastic variable governed by vehicle actuation on the side street.

In speaking of input and output, we do not refer to input and output to the intersection, but only to the congestion at the intersection. Thus, when there is no queue, a car may come into the intersection and pass out of it, leaving both input and output zero. It is essential to specify under what circumstances input is possible. It might be assumed that S is as great as the speed of arriving vehicles, and so input would be zero as soon as the last queued vehicle began to move, provided *that* vehicle cleared the intersection in a single green phase. Such an assumption seems to conflict with our experience and certainly leads to an unwieldy formulation. Consequently we will assume continued input in every circumstance except zero queue length.

2. FIXED CYCLE LIGHTS

Suppose β , T and N are constants. We will have occasion to use the Poisson probability, which will be abbreviated

$$P(u; \Delta) = e^{-\Delta} \Delta^u / u! \quad (u = 0, 1, 2, \dots),$$

and the Borel-Tanner probability [cf. Borel (1942), Tanner (1953), Haight & Breuer (1960)], which will be abbreviated

$$R(u; r, \rho) = A(u, r) e^{-\rho u} \rho^{u-r} \quad (u = r, r+1, \dots),$$

where

$$A(u, r) = \frac{r}{(u-r)!} u^{u-r-1}.$$

This represents the probability that exactly u members of a queue will be served before the queue first becomes empty, beginning with r members and with traffic intensity ρ , assuming Poisson input and regular discharge. In our application, during the green phase, ρ always equals λT [approximately $\alpha\lambda/N$], and will be omitted, writing $R(u; r)$.

The purpose of this paper is to compute the probability of Z cars being in the queue at the beginning of a red phase, when there were X cars in the queue at the beginning of the preceding green phase. Z will be called the overflow into the red phase. If $X > N$, the overflow must be non-zero, and will depend only on the input during the green phase. Letting

$$f(z; x) = \Pr[Z = z \mid X = x],$$

we have

$$f(z; x) = P(z - x + N; \alpha\lambda) \quad (x > N). \quad (1)$$

When $x \leq N$, $z = 0$, the formula is also fairly simple, and can be written in terms of cumulative Borel-Tanner probabilities as follows:

$$f(0; x) = \sum_{j=x}^N R(j; x) \quad (x \leq N). \quad (2)$$

In the remaining cases ($z > 0$, $x \leq N$), consider, for fixed x , the event that the overflow would be z , and that no vehicles would arrive after the light turned red in the period that it would take this overflow to clear. The probability of this is $f(z; x)e^{-\rho z}$, and it may be regarded as the probability that the queue would first vanish after $N + z$ vehicles and that the overflow was z . The unconditional probability of the queue first clearing after $N + z$ vehicles is of course $R(N + z; x)$, and therefore the difference $R(N + z; x) - f(z; x)e^{-\rho z}$ is the sum from $j = 1$ to $j = z - 1$ of the probability that the queue would first vanish after $N + z$ vehicles and that the overflow is j , that is

$$\sum_{j=1}^{z-1} f(j; x) R(z; j),$$

from which we obtain

$$f(z; x) = e^{\rho z} [R(N + z; x) - \sum_{j=1}^{z-1} R(z; j) f(j; x)] \quad (z > 0, x \leq N). \quad (3)$$

Equations (1), (2) and (3) taken together give a formula for $f(z; x)$ in terms of the preceding values $f(j; x)$, $j = 1, 2, \dots, z - 1$, for various values of z and x . In conjunction with these expressions, it is necessary to use the following conventions:

- A. $P(u; \Delta) = 0$ for $u < 0$,
- B. $R(u; r) = 0$ for $u < r$,
- C. $R(u; 0) = 0$ for $u > 0$
 $= 1$ for $u = 0$.

Then it will be noted that $f(z; x) = 0$ whenever $z + N < x$, and this is in fact an impossible transition. The first two values of (3) are

$$\begin{aligned} f(1; x) &= e^{\rho} R(N + 1; x), \\ f(2; x) &= e^{2\rho} [R(N + 2; x) - e^{-\rho} R(N + 1; x)]. \end{aligned}$$

It is also possible to write each $f(z; x)$ explicitly in terms of the Borel-Tanner coefficients $A(z, x)$. If we define $B(z, x)$ by means of the $(z-x)$ th order determinant

$$B(z, x) = (-1)^{z+x} \begin{vmatrix} A(z, z-1) & 1 & 0 & \dots & 0 \\ A(z, z-2) & A(z-1, z-2) & 1 & \dots & 0 \\ A(z, z-3) & A(z-1, z-3) & A(z-2, z-3) & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & 1 \\ A(z, x) & A(z-1, x) & A(z-2, x) & \dots & A(x+1, x) \end{vmatrix} \quad (4)$$

for $z > x$ and $B(z, z) = 1$, then we can write $f(z; x)$ in the form

$$\begin{aligned} f(z; x) &= \sum_{j=1}^z e^{\rho^j} \rho^{z-j} B(z; j) R(N+j; x) \\ &= e^{-\rho^N} \rho^{N+z-x} \sum_{j=1}^z B(z, j) A(N+j, x) \quad (z > 0, x \leq N). \end{aligned} \quad (5)$$

With a table of $A(z, x)$, the coefficients $B(z, x)$ can be most conveniently calculated from the formula

$$B(z, x) = - \sum_{j=x}^{z-1} A(z, j) B(j, x). \quad (6)$$

The first few values of $A(z, x)$ and $B(z, x)$ are given below; to avoid awkward fractions, each value is multiplied by $(z-1)!$

$(z-1)! A(z, x)$

$\begin{smallmatrix} x \\ z \end{smallmatrix}$	1	2	3	4	5	6	7
1	1	—	—	—	—	—	—
2	1	1	—	—	—	—	—
3	3	4	2	—	—	—	—
4	16	24	18	6	—	—	—
5	125	200	180	96	24	—	—
6	1296	2160	2160	1440	600	120	—
7	16807	28812	30870	23520	12600	4320	720

$(z-1)! B(z, x)$

$\begin{smallmatrix} x \\ z \end{smallmatrix}$	1	2	3	4	5	6	7
1	1	—	—	—	—	—	—
2	-1	1	—	—	—	—	—
3	1	-4	2	—	—	—	—
4	-1	12	-18	—	—	—	—
5	1	-32	108	6	—	—	—
6	-1	80	-540	-96	24	—	—
7	1	-192	2430	960	-600	120	—
				-7680	9000	-4320	720

To show that, for fixed x , $\Sigma f(z; x) = 1$, consider

$$\sum_{j=1}^{\infty} e^{-\rho j} f(j; x) = \sum_{j=1}^{\infty} R(N+j; x) - \sum_{j=1}^{\infty} f(j; x) (1 - e^{-\rho j}),$$

which is equivalent to
$$\sum_{j=1}^{\infty} f(j; x) = \sum_{j=1}^{\infty} R(N+j; x).$$

Combining this with (2) gives the value unity provided $x \leq N$; in the opposite case the result is clear from (1).

During the red phase, the transition probabilities are simple Poisson expressions. In terms of these and $f(z; x)$, we can write equations involving other probabilities. For example, if p_{xy} is the probability of a transition from x cars at the beginning of green to y cars at the beginning of the following green, we have

$$\left. \begin{aligned} p_{xy} &= P(y-x+N; (\alpha+\beta)\lambda) \quad (x > N) \\ &= P(y; \lambda\beta) \sum_{j=x}^N R(j; x) + \sum_{z=1}^y P(y-z; \lambda\beta) f(z; x) \quad (x \leq N). \end{aligned} \right\} \quad (7)$$

Also, letting
$$\begin{aligned} \pi_n &= \Pr(n \text{ cars waiting at beginning of green} \mid t = \infty), \\ \sigma_n &= \Pr(n \text{ cars waiting at beginning of red} \mid t = \infty), \end{aligned}$$

we have (if these limiting probabilities exist)

$$\sigma_n = \sum_{j=0}^{n+N} f(n; j) \pi_j \quad (8)$$

and
$$\pi_n = \sum_{j=0}^n P(n-j; \lambda\beta) \sigma_j. \quad (9)$$

3. SEMI-ACTUATED LIGHTS

In this section we consider the traffic along a main street into a signalized intersection governed by a light (possibly) actuated by side street traffic. When the light turns green for the main street, it must remain green for a minimum fixed period (the main street minimum). After the expiration of this time the light will change whenever a car arrives on the side street. Then the light will be red on the main street for a fixed period (side street initial), and a further fixed period (side street extension). If any side street vehicles arrive during the first side street extension the main street light remains red for a further side street extension, measured from the time of arrival of the car. This process continues as long as cars arrive on the side street, or until a maximum value (side street maximum) is reached, when it turns green again. We will assume Poisson arrivals on both streets, and use the following notation:

- λ = main street input parameter,
- μ = side street input parameter,
- a = length of side street initial,
- b = length of side street extension,
- N = number of cars able to clear during main street minimum,
- α = length of main street initial,
- A = length of side street maximum.

These quantities are fixed constants of the system. In addition, we need to define certain stochastic variables which will be used to characterize the system.

β = length of main street red. Defined over $a + b \leq \beta \leq A$, with density function $g(u)$,

n = number of cars able to pass through a main street green period. Defined over $N, N + 1, N + 2, \dots$, with probability distribution q_n . For simplicity we will call n the number of 'slots' provided by the green phase.

Once explicit forms are found for $g(u)$ and q_n , the problem of determining the various distributions of § 2 is solved. For example, an equation like (3) would only need to be modified to take into account the stochastic nature of the number of slots:

$$f(z; x) = \sum_{n=N}^{\infty} e^{\rho z} [R(n+z; x) - \sum_{j=1}^{z-1} R(z; j) f(j; x)] q_n.$$

In case the side street maximum is infinite, both of these distributions are easy to obtain. The red phase length distribution is given in a slightly different context by Raff (1951), and was apparently discovered by Garwood (1940). In the notation defined above, it is

$$g(u) = \mu e^{-\mu b} \sum_{i=0}^{h-1} (-e^{-\mu b})^i \left\{ \frac{[\mu(u - b(1+i))]^{i-1}}{(i-1)!} + \frac{[\mu(u - b(1+i))]^i}{i!} \right\},$$

$$hb \leq u \leq (h+1)b \quad (h = 1, 2, \dots).$$

The slot distribution can be calculated directly from the Poisson probabilities and turns out to be

$$q_n = 1 - e^{-\alpha\mu} \quad (n = N)$$

$$= (1 - e^{-\mu T}) e^{-\mu T(n-1)} \quad (n > N),$$

where $\alpha = NT$.

REFERENCES

- BOREL, E. (1942). Sur l'emploi du théorème de Bernoulli pour faciliter le calcul d'un infinité de coefficients. Application au problème de l'attente à un guichet. *C.R. Acad. Sci., Paris*, **214**, 452-6.
- GARWOOD, F. (1940). An application of the theory of probability to the operation of vehicular-controlled traffic signals. *J. R. Statist. Soc. (Suppl.)*, **7**, 65-77.
- HAIGHT, F. A. & BREUER, M. A. (1960). The Borel-Tanner distribution. (To appear.)
- RAFF, M. S. (1951). The distribution of blocks in an uncongested stream of automobile traffic. *J. Amer. Statist. Ass.* **46**, 114-23.
- TANNER, J. C. (1953). A problem of interference between two queues. *Biometrika*, **38**, 58-69.

PARTIAL TESTS FOR PARTIAL TAUS*

BY LEO A. GOODMAN

University of Chicago

1. INTRODUCTION AND SUMMARY

In his 1942 paper, Kendall introduced and discussed the partial rank correlation coefficient τ (τ). Since then, the difficulties involved in developing tests of significance for partial τ have been discussed by Kendall (1948) and the sampling distribution of partial τ has been studied by Hoeffding (1948) and Moran (1951). This work has been briefly summarized by Kendall (1955) as follows: 'No tests of significance are yet known for partial τ Hoeffding (1948) gives a rather complicated expression for the variance of partial τ in the limiting case when n [the sample size] is large....Moran (1951) has considered partial τ without reaching any clear conclusions other than that the distributional problem is a very complex one.' In the present paper, tests of significance will be presented for a number of different partial correlation coefficients that are closely related to partial τ . These tests will be valid when the sample size is large and when certain additional assumptions are made about the situations under consideration (see § 2).

The partial correlation coefficients introduced herein are computed from ranked data, as is partial τ , and they measure various aspects of what is usually meant by partial association. For the situations considered in the present paper, some justification can be found for the use of a weighted average of the partial correlation coefficients introduced herein as an over-all measure of partial association; less justification can be found for the use of partial τ , which is also a summary measure, in a certain general sense, related to these partial correlation coefficients. For some justification for the use of partial τ in situations that differ somewhat from those considered herein, the reader is referred to Kendall (1942), and Hoeffding (1948).

The partial correlation coefficients discussed herein are related to, but differ from, the generalized partial correlation coefficients introduced by Somers (1959). The methods developed in the present paper could, no doubt, be generalized in order to obtain tests of significance for various kinds of partial correlation coefficients. For the situations considered in the present paper, the use of the particular partial correlation coefficients discussed herein has some justification; these coefficients can be given simple probabilistic and operational interpretations, and the tests of significance developed for these coefficients are quite easy to apply.

2. THE MEASUREMENT OF PARTIAL ASSOCIATION

Partial association usually refers to the study of whether an association or correlation of quality Q with quality R is really due to the associations of each with a third quality P . Theories of partial association usually attempt to decide the matter by the consideration of subpopulations in which the variation of P is eliminated (see, for example, Kendall,

* Research carried out at the Statistical Research Center, University of Chicago, under sponsorship of the Statistics Branch, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

1955). The partial association between Q and R , when the variation of P is, in a certain sense, eliminated, will be studied herein in situations where the statistical analysis is to be based on the rank orders of the observed values of Q , of R , and of P .

We shall follow the usual practice of arranging the ranks of the observed values of P in their natural order $1, 2, 3, \dots, n$, and of assuming that no ties are present among these observed values (see Kendall, 1955). Let P_i denote the rank of the i th observation (in their natural order) with regard to quality P , so that $P_i = i$ for $i = 1, 2, \dots, n$. We shall assume herein that the conditional joint distribution

$$\Pr\{Q < q, R < r \mid P_i\} = F(q, r \mid P_i)$$

of the variables Q and R , when P_i is given, is continuous so that the problem of ties among the observed values of Q and R can also be ignored. Thus, we assume that Q and R are continuous random variables. Only the rank orders of the actual observations of Q and R (and not the actual observations themselves) will be used in the subsequent statistical analysis. The symbols Q_i and R_i will denote the ranks of the actual observations $Q^{(i)}$ and $R^{(i)}$, respectively, associated with a given P_i .

If $Q^{(i)}$ and $R^{(i)}$ are statistically independent of each other when P_i is given, then

$$\Pr\{Q^{(i)} < q, R^{(i)} < r \mid P_i\} = F(q, r \mid P_i)$$

is the product of the conditional distribution $G(q \mid P_i)$ of Q given P_i and the conditional distribution $H(r \mid P_i)$ of R given P_i ; i.e.

$$F(q, r \mid P_i) = G(q \mid P_i) H(r \mid P_i) \quad (i = 1, 2, \dots, n).$$

In this case, by adapting the usual meaning of partial association we find that the partial association between Q and R is zero when the variation of P is eliminated in a certain sense, but the expected 'association between the agreements of [rankings] Q with P and those of R with P ' will in general not be zero. To see that the expected association between the agreements of rankings Q and R with P will in general not be equal to zero even when

$$F(q, r \mid P_i) = G(q \mid P_i) H(r \mid P_i) \quad (i = 1, 2, \dots, n),$$

consider the situation where the following two conditions are true:

(I) the n pairs $(Q^{(i)}, R^{(i)})$ are statistically independent observations (the conditional distribution $F(q, r \mid P_i)$ of each pair may depend on P_i).

(II) $\Pr\{Q^{(i)} > Q^{(j)} \mid P_i, P_j\} = b_{i-j}, \quad \Pr\{R^{(i)} > R^{(j)} \mid P_i, P_j\} = c_{i-j}.$

(Condition (II) attaches somewhat more significance to the ranks of quality P than is usually done. This condition will be satisfied if, for example, the conditional distributions $G(q \mid P_i)$, where $P_i = 1, 2, \dots, n$, differ only in location, the location parameter is a linear function of P_i , and a similar statement holds true for $H(r \mid P_i)$; this will often be the case when $Q^{(i)}$ and $R^{(i)}$ are linearly related to P_i . It should be remembered that linearity also plays an important role in the usual theories of partial correlation for quantitative variables (see, for example, Cramér, 1946). For a description of some situations where conditions (I) and (II) can be assumed to hold true, the reader is referred to Goodman & Grunfeld, 1959.) Assuming conditions (I) and (II), we see that, when $F(q, r \mid P_i) = G(q \mid P_i) H(r \mid P_i)$, the following condition holds true:

(III) $\Pr\{Q^{(i)} > Q^{(j)}, R^{(i)} > R^{(j)} \mid P_i, P_j\} = b_{i-j} c_{i-j}.$

Considering the $\frac{1}{2}n(n-1)$ pairs (Q_i, R_i, P_i) , (Q_j, R_j, P_j) , where $i > j$, the expected proportion of these pairs that are such that $Q_i > Q_j$ and $R_i > R_j$ (given that $P_i > P_j$) is $a = \sum_{i>j} b_{i-j} c_{i-j} 2/n(n-1)$, while the expected proportion of these pairs that are such that $Q_i > Q_j$ is $b = \sum_{i>j} b_{i-j} 2/n(n-1)$, and the expected proportion that are such that $R_i > R_j$ is $c = \sum_{i>j} c_{i-j} 2/n(n-1)$. If the expected association between the agreements of rankings Q with P and those of R with P were zero, then the expected proportion of these pairs that are such that $Q_i > Q_j$ and $R_i > R_j$ would be bc . Since a is in general not equal to bc , the expected association between the agreements of rankings Q with P and those of R with P will differ from zero even when the partial association (in the usual sense) between Q and R , given P_i , is zero. Since $\tau_{QR.P}$ measures the intensity of association between the agreements of Q with P and those of R with P (see Kendall, 1955), justification of its use as a measure of partial association between Q and R , given P , is doubtful for situations of the kind considered herein. It is only in the special case where $\sum_{i>j} b_{i-j}[c_{i-j} - c] = 0$ that the expected association between agreements is zero when the partial association is zero. Thus, $\tau_{QR.P}$ does not measure the agreement between Q and R independently of the influence of P_i except perhaps in situations where the 'covariance' between b_{i-j} and c_{i-j} is zero. The study of the partial association between Q and R , given P , is usually of interest in situations where the 'covariance' between b_{i-j} and c_{i-j} is not zero.

Assuming conditions (I), (II) and (III) described in the preceding paragraph, we see that for the $(n-1)$ pairs (Q_i, R_i, P_i) , (Q_j, R_j, P_j) , where $i = j+1$, the expected proportion of these pairs that are such that $Q_i > Q_j$ and $R_i > R_j$ (given that $P_i > P_j$) is $a_1 = b_1 c_1$, while the expected proportion of these pairs that are such that $Q_i > Q_j$ is b_1 , and the expected proportion of those that are such that $R_i > R_j$ is c_1 . Since $a_1 = b_1 c_1$, the expected association between the agreements of rankings Q with P and those of R with P , when the $(n-1)$ pairs of observations where $i = j+1$ are compared, will be zero when the partial association (in the usual sense) between Q and R , given P_i , is zero. Thus, there is some justification for the use of a measure of the association τ_1 between the agreements of rankings Q with P and those of R with P , as observed in the fourfold table giving the joint distribution of the agreements for the $(n-1)$ pairs of observations where $i = j+1$, as a measure of an aspect of the partial association between Q and R , given P , in the situation considered here. The measure of association τ_1 based on the fourfold table for the $(n-1)$ pairs of observations where $i = j+1$ could be computed in an analogous fashion to the usual computation of partial τ based on the fourfold table for the $\frac{1}{2}n(n-1)$ pairs of observations where $i > j$ (see Kendall 1948); it could be given a probabilistic and operational interpretation similar to that given to such measures of association for fourfold tables and also analogous to, but somewhat different from, the usual interpretation for τ (see Kendall (1955), Kruskal (1958), Goodman & Kruskal (1959)).

By an approach similar to that presented above, we see that a measure of the association between Q and R , given P , can be based on a measure of the association between the agreements of rankings Q with P and those of R with P , as observed in the fourfold table giving the joint distribution of the agreements for the $(n-2)$ pairs of observations where $i = j+2$. In general, the fourfold table giving the joint distribution of the agreements for the $(n-k)$ pairs of observations where $i = j+k$ (where k is a fixed constant), can be used to obtain a measure of the association τ_k between agreements that will measure an aspect of the partial association between Q and R , given P . The symbol k can denote any

fixed constant between 1 and $n-1$, but when k is close to $n-1$ the frequencies appearing in the fourfold table will be too small. For each k , a different fourfold table will be obtained, and thus a different measure of association τ_k will also be obtained. Each of these measures can be given a probabilistic and operational interpretation.

The fourfold table described by Kendall (1948), which is used to define partial τ , is the table obtained when the corresponding entries in the separate fourfold tables, computed for each fixed k , are added together for $k = 1, 2, \dots, n-1$; i.e. this table gives the joint distribution of the agreements for the $\frac{1}{2}n(n-1)$ pairs of observations where $i = j+k$ for $k > 0$, and partial τ is a measure of the association between agreements in this table. Since the table obtained when the corresponding entries of a number of separate tables are added together need not indicate independence between the two qualities described by the table, even when there is complete independence in each of the separate fourfold tables, partial τ may differ from zero even when the measure of association τ_k computed for each of the separate tables (for $k = 1, 2, \dots$) is zero. Some justification can be found for basing a measure of partial association between Q and R , given P , on a weighted average of the measures of association τ_k obtained from the separate fourfold tables ($k = 1, 2, \dots$), but less justification can be found, in the situation considered here, for the use of partial τ , a measure of association obtained from a fourfold table, which is a weighted average, in a certain sense, of the separate tables (except in the special case where the 'covariance' between b_{i-j} and c_{i-j} is zero).

For the situation considered herein, the difference between the ranks P_i and P_j is meaningful. This meaningfulness leads in a natural way to the introduction of the separate fourfold tables and the corresponding measures of association τ_k . Although no simple tests of significance are yet known for τ , it will be possible to develop such tests for the τ_k in the present situation (see § 3). For situations where the difference between the ranks P_i and P_j is not meaningful, some justification can be given for the use of partial τ (see Kendall (1942) and Hoeffding (1948)); less justification can be found for the use of τ_k in these cases.

For some further discussion of the relation between ordinal measures of association, such as τ or partial τ , and measures of association computed for cross-classification tables, such as the fourfold tables considered here, the reader is referred to Kendall (1948), Kruskal (1958), Goodman & Kruskal (1959). A discussion of measures of partial association computed for cross-classification tables will be found in Goodman & Kruskal (1954).

3. TESTS OF SIGNIFICANCE

Let us first consider the fourfold table giving the joint distribution of the agreements of rankings Q with P and those of R with P , for the $(n-1)$ pairs (Q_i, R_i, P_i) , (Q_j, R_j, P_j) , where $i = j+1$.

It is easy to see that the usual χ^2 test of independence for fourfold tables can not be used in this case because the $(n-1)$ pairs of observations, where $i = j+1$ and $j = 1, 2, \dots, n-1$, are not statistically independent. However, assuming conditions (I) and (II) described in the preceding section, the $\frac{1}{2}(n-1)$ pairs where $i = j+1$ and $j = 1, 3, 5, \dots, n-2$ (in the case where n is odd) are statistically independent, given the P_i , and so are the $\frac{1}{2}(n-1)$ pairs where $i = j+1$ and $j = 2, 4, 6, \dots, n-1$. Thus, if condition (III) of the preceding section is also assumed, then the usual large sample theory for tests of independence for fourfold tables can be applied in order to prove that the asymptotic distribution of

$$\tau_{11}\sqrt{n_1} = (\hat{a}_{11} - \hat{b}_{11}\hat{c}_{11})/\sqrt{\{n_1/\hat{b}_{11}\hat{c}_{11}(1-\hat{b}_{11})(1-\hat{c}_{11})\}}$$

is normal with zero mean and unit variance, where $\hat{a}_{11}$ is the proportion of the $\frac{1}{2}(n-1) = n_1$ pairs ($i = j+1, j = 1, 3, 5, \dots, n-2$) that are such that $Q_i > Q_j$ and $R_i > R_j$, $\hat{b}_{11}$ is the proportion of these n_1 pairs that are such that $Q_i > Q_j$, $\hat{c}_{11}$ is the proportion of these n_1 pairs that are such that $R_i > R_j$ (see, for example, Pearson & Hartley, 1954, p. 71). Similarly, the asymptotic distribution of

$$\tau_{12} \sqrt{n_1} = (\hat{a}_{12} - \hat{b}_{12} \hat{c}_{12}) / \sqrt{\{n_1 / \hat{b}_{12} \hat{c}_{12} (1 - \hat{b}_{12}) (1 - \hat{c}_{12})\}}$$

is normal with zero mean and unit variance, where $\hat{a}_{12}$ is the proportion of the n_1 pairs ($i = j+1, j = 2, 4, \dots, n-1$) that are such that $Q_i > Q_j$ and $R_i > R_j$, $\hat{b}_{12}$ is the proportion of these n_1 pairs that are such that $Q_i > Q_j$, $\hat{c}_{12}$ is the proportion of these n_1 pairs that are such that $R_i > R_j$. Thus, a test of the null hypothesis that condition (III) is true can be based on either $\tau_{11} \sqrt{n_1}$ or on $\tau_{12} \sqrt{n_1}$. Since these two statistics are correlated, they cannot be combined in the most direct fashion to form a single test of significance. However, it can be seen that the covariance between $(\hat{a}_{11} - \hat{b}_{11} \hat{c}_{11}) \sqrt{n_1}$ and $(\hat{a}_{12} - \hat{b}_{12} \hat{c}_{12}) \sqrt{n_1}$ is asymptotically $2f_1 g_1$, where

$$f_1 = d_1 - b_1^2, \quad d_1 = \Pr\{Q_{i+1} > Q_i > Q_{i-1}\}, \\ g_1 = e_1 - c_1^2 \quad \text{and} \quad e_1 = \Pr\{R_{i+1} > R_i > R_{i-1}\},$$

assuming that d_1 and e_1 are constant and do not depend on i (see Goodman & Grunfeld, 1959). This result can be generalized to situations where d_1 and e_1 are not constant, but the generalization will not be presented here (see Hoeffding & Robbins, 1948, for some results related to this generalization). Using this result concerning the asymptotic covariance $2f_1 g_1$, we find that the distribution of

$$z_1 = (\hat{a}_1 - \hat{b}_1 \hat{c}_1) / \sqrt{\{(n-1) / [\hat{b}_1 \hat{c}_1 (1 - \hat{b}_1) (1 - \hat{c}_1) + 2f_1 g_1]\}}$$

is normal with zero mean and unit variance, where $\hat{a}_1 = \frac{1}{2}(\hat{a}_{11} + \hat{a}_{12})$ is the proportion of the $(n-1)$ pairs ($i = j+1, j = 1, 2, 3, \dots, n-1$) that are such that $Q_i > Q_j$ and $R_i > R_j$,

$$\hat{b}_1 = \frac{1}{2}(\hat{b}_{11} + \hat{b}_{12}), \quad \hat{c}_1 = \frac{1}{2}(\hat{c}_{11} + \hat{c}_{12}), \quad \hat{f}_1 = \hat{d}_1 - \hat{b}_1^2, \quad \hat{g}_1 = \hat{e}_1 - \hat{c}_1^2,$$

$\hat{d}_1$ is the proportion of the $(n-2)$ triplets $\{Q_{i+1}, Q_i, Q_{i-1}\}$ that are such that

$$Q_{i+1} > Q_i > Q_{i-1},$$

and $\hat{e}_1$ is the proportion of the $(n-2)$ triplets $\{R_{i+1}, R_i, R_{i-1}\}$ that are such that

$$R_{i+1} > R_i > R_{i-1}$$

(see Goodman & Grunfeld, 1959, for a detailed proof and also for some applications of this result). Since the measure of association τ_1 between the agreements of rankings Q with P and those of R with P , as observed in the fourfold table giving the joint distribution of the agreements for the $(n-1)$ pairs of observations where $i = j+1$ is simply

$$\tau_1 = (\hat{a}_1 - \hat{b}_1 \hat{c}_1) / \sqrt{\{\hat{b}_1 \hat{c}_1 (1 - \hat{b}_1) (1 - \hat{c}_1)\}}$$

we see that the asymptotic distribution of $z_1 = \tau_1 \sqrt{\{(n-1) / \hat{h}_1\}}$ is normal with zero mean and unit variance, where

$$\hat{h}_1 = 1 + [2\hat{f}_1 \hat{g}_1 / \hat{b}_1 \hat{c}_1 (1 - \hat{b}_1) (1 - \hat{c}_1)].$$

Thus, a test of the null hypothesis that condition (III) is true can be based on τ_1 or, more precisely, on $\tau_1 \sqrt{\{(n-1) / \hat{h}_1\}}$. For the situation considered here, this test of significance based on τ_1 provides a partial test of the hypothesis that the partial association between Q and R , given P_i , is zero.

Let us now consider the fourfold table giving the joint distribution of the agreements of rankings Q with P and those of R with P , for the $(n-2)$ pairs (Q_i, R_i, P_i) , (Q_j, R_j, P_j) , where $i = j+2$. The usual χ^2 test of independence for fourfold tables cannot be applied in this case. However, assuming conditions (I) and (II), the $\frac{1}{2}(n-2) = n_2$ pairs ($i = j+2, j = 1, 2, 5, 6, 9, 10, 13, 14, \dots$, in the case where n_2 is even) are statistically independent, given the P_i , and so are the $\frac{1}{2}(n-2)$ pairs ($i = j+2, j = 3, 4, 7, 8, 11, 12, \dots$). Thus, if condition (III) is also assumed, then the large sample theory for tests of independence for fourfold tables can be applied in order to prove that the asymptotic distribution of

$$\tau_{21}\sqrt{n_2} = (\hat{a}_{21} - \hat{b}_{21}\hat{c}_{21})\sqrt{\{n_2/\hat{b}_{21}\hat{c}_{21}(1-\hat{b}_{21})(1-\hat{c}_{21})\}}$$

is normal with zero mean and unit variance, where $\hat{a}_{21}$ is the proportion of the n_2 pairs ($i = j+2, j = 1, 2, 5, 6, 9, 10, \dots$) that are such that $Q_i > Q_j$ and $R_i > R_j$, $\hat{b}_{21}$ is the proportion of these n_2 pairs that are such that $Q_i > Q_j$, $\hat{c}_{21}$ is the proportion of these n_2 pairs that are such that $R_i > R_j$. Similarly, the asymptotic distribution of

$$\tau_{22}\sqrt{n_2} = (\hat{a}_{22} - \hat{b}_{22}\hat{c}_{22})\sqrt{\{n_2/\hat{b}_{22}\hat{c}_{22}(1-\hat{b}_{22})(1-\hat{c}_{22})\}}$$

is normal with zero mean and unit variance, where $\hat{a}_{22}$, $\hat{b}_{22}$ and $\hat{c}_{22}$ are defined in an analogous way for the n_2 pairs ($i = j+2, j = 3, 4, 7, 8, 11, 12, \dots$). The null hypothesis that condition (III) is true can be based on either $\tau_{21}\sqrt{n_2}$ or on $\tau_{22}\sqrt{n_2}$. It can be seen that the covariance between $(\hat{a}_{21} - \hat{b}_{21}\hat{c}_{21})\sqrt{n_2}$ and $(\hat{a}_{22} - \hat{b}_{22}\hat{c}_{22})\sqrt{n_2}$ is asymptotically $2f_2g_2$, where

$$f_2 = d_2 - b_2^2, \quad d_2 = \Pr\{Q_{i+2} > Q_i > Q_{i-2}\},$$

$$g_2 = e_2 - c_2^2 \quad \text{and} \quad e_2 = \Pr\{R_{i+2} > R_i > R_{i-2}\},$$

assuming that d_2 and e_2 are not functions of i . (This result can be generalized somewhat, but this generalization will not be presented here.) Also, the distribution of

$$z_2 = (\hat{a}_2 - \hat{b}_2\hat{c}_2)\sqrt{\{(n-2)/[\hat{b}_2\hat{c}_2(1-\hat{b}_2)(1-\hat{c}_2) + 2\hat{f}_2\hat{g}_2]\}}$$

can be seen to be normal with zero mean and unit variance, where $\hat{a}_2 = \frac{1}{2}(\hat{a}_{21} + \hat{a}_{22})$ is the proportion of the $n-2$ pairs ($i = j+2, j = 1, 2, \dots, n-2$) that are such that $Q_i > Q_j$ and $R_i > R_j$,

$$\hat{b}_2 = \frac{1}{2}(\hat{b}_{21} + \hat{b}_{22}), \quad \hat{c}_2 = \frac{1}{2}(\hat{c}_{21} + \hat{c}_{22}), \quad \hat{f}_2 = \hat{d}_2 - \hat{b}_2^2, \quad \hat{g}_2 = \hat{e}_2 - \hat{c}_2^2,$$

$\hat{d}_2$ is the proportion of the $(n-4)$ triplets $\{Q_{i+2}, Q_i, Q_{i-2}\}$ that are such that $Q_{i+2} > Q_i > Q_{i-2}$, and $\hat{e}_2$ is the proportion of the $(n-4)$ triplets that are such that $R_{i+2} > R_i > R_{i-2}$. Since the measure of association τ_2 between the agreements of rankings Q with P and those of R with P , as observed in the fourfold table giving the joint distribution of the agreements for the $(n-2)$ pairs of observations where $i = j+2$ is simply

$$\tau_2 = (\hat{a}_2 - \hat{b}_2\hat{c}_2)/\sqrt{\{\hat{b}_2\hat{c}_2(1-\hat{b}_2)(1-\hat{c}_2)\}},$$

we see that the asymptotic distribution of $z_2 = \tau_2\sqrt{\{(n-2)/\hat{h}_2\}}$ is normal with zero mean and unit variance, where

$$\hat{h}_2 = 1 + [2\hat{f}_2\hat{g}_2/\hat{b}_2\hat{c}_2(1-\hat{b}_2)(1-\hat{c}_2)].$$

Thus, a test of the null hypothesis that condition (III) is true can be based on τ_2 (i.e. on $\tau_2\sqrt{\{(n-2)/\hat{h}_2\}}$), which provides a partial test of significance of the partial association between Q and R , given P_i .

Let us now consider the general case where the fourfold table giving the joint distribution of the agreements for the $(n-k)$ pairs of observations $(Q_i, R_i, P_i), (Q_j, R_j, P_j)$, where $i = j+k$, is of interest, where k is a fixed constant. The special cases where $k = 1$ and where $k = 2$ were discussed above. The measure of association τ_k between the agreements of rankings Q with P and those of R with P , as observed in this fourfold table for the $(n-k)$ pairs of observations, is simply

$$\tau_k = (\hat{a}_k - \hat{b}_k \hat{c}_k) \sqrt{\{\hat{b}_k \hat{c}_k (1 - \hat{b}_k) (1 - \hat{c}_k)\}},$$

where $\hat{a}_k$ is the proportion of the $(n-k)$ pairs that are such that $Q_i > Q_j$ and $R_i > R_j$, $\hat{b}_k$ is the proportion of these $(n-k)$ pairs that are such that $Q_i > Q_j$, and $\hat{c}_k$ is the proportion of the $(n-k)$ pairs that are such that $R_i > R_j$. We shall assume that the following condition holds true:

(IIa) $d_k = \Pr\{Q_{i+k} > Q_i > Q_{i-k}\}$ and $e_k = \Pr\{R_{i+k} > R_i > R_{i-k}\}$ are not functions of i (this condition can be weakened in order to obtain somewhat more general results (see related results by Hoeffding & Robbins, 1948)).

By an argument similar to that presented earlier herein, it can be seen that, under conditions (I), (II), (IIa) and (III), the asymptotic distribution ($n \rightarrow \infty$) of $\tau_k \sqrt{\{(n-k)/\hat{h}_k\}}$ is normal with zero mean and unit variance, when

$$\hat{h}_k = 1 + [2\hat{f}_k \hat{g}_k / \hat{b}_k \hat{c}_k (1 - \hat{b}_k) (1 - \hat{c}_k)], \quad \hat{f}_k = \hat{d}_k - \hat{b}_k^2, \quad \hat{g}_k = \hat{e}_k - \hat{c}_k^2,$$

$\hat{d}_k$ is the proportion of the $(n-2k)$ triplets (Q_{i+k}, Q_i, Q_{i-k}) that are such that $Q_{i+k} > Q_i > Q_{i-k}$, and $\hat{e}_k$ is the proportion of the $(n-2k)$ triplets $\{R_{i+k}, R_i, R_{i-k}\}$ that are such that

$$R_{i+k} > R_i > R_{i-k}$$

The quantity $\hat{f}_k$ is an estimate of the 'serial covariance' of order k in the sequence of random variables $S(Q_{k+1} - Q_1), S(Q_{k+2} - Q_2), S(Q_{k+3} - Q_3), \dots$, and $\hat{g}_k$ is an estimate of the 'serial covariance' of order k in the sequence $S(R_{k+1} - R_1), S(R_{k+2} - R_2), S(R_{k+3} - R_3), \dots$, where $S(X) = 1$ when $X > 0$ and $S(X) = 0$ when $X \leq 0$. A test of the null hypothesis that condition (III) is true can be based on τ_k (i.e. on $\tau_k \sqrt{\{(n-k)/\hat{h}_k\}}$), which provides a partial test of significance of the partial association between Q and R , given P_i .

The tests here described have been applied, in the special case where $k = 1$, by Goodman & Grunfeld (1959). The reader is referred to this article for a more detailed discussion of the special case, and also for some numerical examples illustrating the application of the tests based on τ_1 to the analysis of comovements in time series. The tests based on τ_k (for $k = 1, 2, \dots$) described herein could also be applied to the analysis of co-movements in time series in a manner similar to the application of τ_1 .

REFERENCES

- CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.
- GOODMAN, L. A. & GRUNFELD, Y. (1959). Some nonparametric tests for comovements between time series. To be submitted to the *J. Amer. Statist. Ass.*
- GOODMAN, L. A. & KRUSKAL, W. H. (1954). Measures of association for cross classifications. *J. Amer. Statist. Ass.* **49**, 732-64.
- GOODMAN, L. A. & KRUSKAL, W. H. (1959). Measures of association for cross classifications. II. Further discussion and references. *J. Amer. Statist. Ass.* **54**, 124-63.
- HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19**, 293-325.
- HOEFFDING, W. & ROBBINS, H. (1948). The central limit theorem for dependent random variables. *Duke Math. J.* **15**, 773-80.

- KENDALL, M. G. (1942). Partial rank correlation. *Biometrika*, **32**, 277-83.
- KENDALL, M. G. (1955). *Rank Correlation Methods*. (Second edition.) London: Charles Griffin and Co.
- KRUSKAL, W. H. (1958). Ordinal measures of association. *J. Amer. Statist. Ass.* **53**, 814-61.
- MORAN, P. A. P. (1951). Partial and multiple rank correlation. *Biometrika*, **38**, 26-32.
- PEARSON, E. S. & HARTLEY, H. O. (1954). *Biometrika Tables for Statisticians*, **1**. Cambridge University Press.
- SOMERS, R. H. (1959). The rank analogue of product-moment partial correlation and regression, with application to manifold, ordered contingency tables. *Biometrika*, **46**, 241-6.

MOMENTS OF ORDER STATISTICS FROM A NORMAL POPULATION*

BY R. C. BOSE† AND SHANTI S. GUPTA‡

University of North Carolina

1. INTRODUCTION AND SUMMARY

Statistics based on ordered observations have been called systematic statistics by Mosteller (1946). They are now being increasingly used in new statistical procedures (Bahadur, 1950; Bechhofer, 1954; Bechhofer, Dunnett & Sobel, 1954; Bechhofer & Sobel, 1956; Godwin, 1949; Gupta, 1956; Mosteller, 1948; Nair, 1948; Paulson, 1949, 1952; Seal, 1955). The present paper deals with the problem of obtaining the moments of $X_{(k)}$, the k th order statistic for a sample of size n from a normal population $N(0, 1)$. This problem has been considered among others by Cole (1951), Godwin (1949), Hastings, Mosteller, Tukey & Winsor (1947), Hojo (1931), Jones (1948), Paulson (1952), Ruben (1954), Teichroew (1956) and Tippett (1925).

It has been shown that $\mu'_t(n, k)$, the t th moment of $X_{(k)}$, can be expressed in terms of lower moments of order $t - 2i$ ($i = 1, 2, \dots, \frac{1}{2}t$ or $\frac{1}{2}(t - 1)$) and the integral

$$\int_{-\infty}^{+\infty} P_{t+1}(x) e^{-\frac{1}{2}(t+1)x^2} dx, \quad (1.1)$$

where $P_{t+1}(x)$ for $t > 0$, is defined by

$$P_{t+1}(x) = k \binom{n}{k} \frac{d^t}{d\Phi^t} [\Phi^{k-1}(1 - \Phi)^{n-k}], \quad (1.2)$$

it being understood that in (1.2), Φ is replaced after differentiation by $\Phi(x)$, the cumulative distribution function (c.d.f.) of $N(0, 1)$. $P_t(x)$ is thus a polynomial of degree $(n - t)$ in $\Phi(x)$ if $t \leq n$ and is zero if $t > n$. Exact values of all odd order moments can be derived when $n \leq 6$. Godwin (1949) and Jones (1948) have given tables of exact moments $\mu'_t(n, k)$ for $t = 1$ and 2. The corresponding tables for $t = 3$ and 4 are provided in this paper. In general the numerical evaluation of the integral (1.1) can be expeditiously done by using the Gauss method of numerical integration (Szego, 1939) based on the zeros and the weight factors of the Hermite-polynomials for which tables have been provided by Salzer, Zucker & Capuano (1952). A sufficiently large number of decimal places must, however, be retained in the computation, to secure an adequate number of significant figures in the final moments about the mean.

2. THE FUNCTION $P_t(n, k, x)$

Let $x_1, x_2, \dots, x_n$ be n independent observations from a normal population $N(0, 1)$ with zero mean and unit variance, and let

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)} \quad (2.1)$$

* This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command. Reproduction in whole or in part is permitted for any purpose of the United States Government.

† Now at the Case Institute of Technology, Cleveland, Ohio.

‡ Now at the Bell Telephone Laboratories, Allentown, Pennsylvania.

be the n ranked observations among $x_1, x_2, \dots, x_n$. Then the cumulative distribution function of $X_{(k)}$, the random variable corresponding to $x_{(k)}$, ($1 \leq k \leq n$), is given by

$$P_0(n, k, x) = \text{Prob} \{X_{(k)} \leq x\} \\ = \frac{C}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^x [\Phi(x)]^{k-1} [1 - \Phi(x)]^{n-k} e^{-\frac{1}{2}x^2} dx, \quad (2.2)$$

where $\Phi(x)$ is defined as
$$\Phi(x) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^x e^{-\frac{1}{2}x^2} dx, \quad (2.3)$$

and C is the constant
$$C = \frac{n!}{(k-1)!(n-k)!}. \quad (2.4)$$

Let us now define the function $P_t(n, k, x)$, which we shall abbreviate to $P_t(x)$ for convenience, by the relation

$$P_{t+1}(x) = (2\pi)^{\frac{1}{2}} e^{\frac{1}{2}x^2} \frac{dP_t}{dx}, \quad (2.5)$$

where $P_0(x)$ is given by (2.2). Then

$$P_1(x) = C[\Phi(x)]^{k-1} [1 - \Phi(x)]^{n-k}. \quad (2.6)$$

It is clear that $P_t(x)$ is a polynomial of degree $n-t$ in $\Phi(x)$ if $1 \leq t \leq n$, and is zero for $t > n$. In fact, we can write

$$P_{t+1}(x) = C \frac{d^t}{d\Phi^t} [\Phi^{k-1}(1-\Phi)^{n-k}], \quad (2.7)$$

where Φ is replaced by $\Phi(x)$ after the differentiation. It follows that for given t, n, k , $\Phi_t(x)$ is a bounded function of x . The functions $P_2(x)$, $P_3(x)$, $P_4(x)$ and $P_5(x)$ are given below explicitly, where Φ is written for $\Phi(x)$.

$$P_2(x) = C\Phi^{k-2}(1-\Phi)^{n-k-1}[(k-1)-(n-1)\Phi], \quad (2.8)$$

$$P_3(x) = C\Phi^{k-3}(1-\Phi)^{n-k-2}[(k-1)(k-2)-2(k-1)(n-2)\Phi+(n-1)(n-2)\Phi^2], \quad (2.9)$$

$$P_4(x) = C\Phi^{k-4}(1-\Phi)^{n-k-3}[(k-1)(k-2)(k-3)-3(k-1)(k-2)(n-3)\Phi \\ + 3(k-1)(n-2)(n-3)\Phi^2 - (n-1)(n-2)(n-3)\Phi^3], \quad (2.10)$$

$$P_5(x) = C\Phi^{k-5}(1-\Phi)^{n-k-4}[(k-1)(k-2)(k-3)(k-4)-4(k-1)(k-2)(k-3)(n-4)\Phi \\ + 6(k-1)(k-2)(n-3)(n-4)\Phi^2 - 4(k-1)(n-2)(n-3)(n-4)\Phi^3 \\ + (n-1)(n-2)(n-3)(n-4)\Phi^4]. \quad (2.11)$$

3. A SYSTEM OF DIFFERENTIAL EQUATIONS SATISFIED BY $P_0(x)$

From (2.5)

$$P_1(x) = (2\pi)^{\frac{1}{2}} e^{\frac{1}{2}x^2} \frac{dP_0}{dx}, \quad (3.1)$$

$$P_2(x) = (2\pi)^{\frac{1}{2}} e^{\frac{1}{2}x^2} \frac{d}{dx} \left[(2\pi)^{\frac{1}{2}} e^{\frac{1}{2}x^2} \frac{dP_0}{dx} \right] \\ = (2\pi) e^{x^2} \left[\frac{d^2 P_0}{dx^2} + x \frac{dP_0}{dx} \right]. \quad (3.2)$$

In general let us assume

$$P_t(x) = (2\pi)^{\frac{1}{2}t} e^{\frac{1}{2}tx^2} \sum_{r=0}^{t-1} g_{r,t}(x) \frac{d^{t-r}P_0}{dy^{t-r}}, \quad (3.3)$$

where

$$g_{0,t}(u) = 0, \quad (3.4)$$

and $g_{r,t}(x)$ is a polynomial in x of the r th degree. Differentiating (3.3) and using (2.5), we have

$$P_{t+1}(x) = (2\pi)^{\frac{1}{2}(t+1)} e^{\frac{1}{2}(t+1)x^2} \sum_{r=0}^{t-1} \left[g_{r,t}(x) \frac{d^{t-r+1}P_0}{dx^{t-r+1}} + \{txg_{r,t}(x) + g'_{r,t}(x)\} \frac{d^{t-r}P_0}{dx^{t-r}} \right]. \quad (3.5)$$

This leads to the recurrence relation

$$g_{r,t+1}(x) = g_{r,t}(x) + \left\{ tx + \frac{d}{dx} \right\} g_{r-1,t}(x), \quad (3.6)$$

where $g_{r,r}(x)$ should be interpreted as zero. This together with (3.4) determines all the polynomials $g_{r,t}(x)$. Starting from

$$g_{0,1}(x) = 1 \quad (3.7)$$

we can successively calculate

$$g_{0,2}(x) = 1, \quad g_{1,2}(x) = x, \quad (3.8)$$

$$g_{0,3}(x) = 1, \quad g_{1,3}(x) = 3x, \quad g_{2,3}(x) = 2x^2 + 1, \quad (3.9)$$

$$g_{0,4}(x) = 1, \quad g_{1,4}(x) = 6x, \quad g_{2,4}(x) = 11x^2 + 4, \quad g_{3,4}(x) = 6x^3 + 7x, \quad (3.10)$$

$$g_{0,5}(x) = 1, \quad g_{1,5}(x) = 10x, \quad g_{2,5}(x) = 35x^2 + 10, \quad g_{3,5}(x) = 50x^3 + 45x, \quad (3.11)$$

Hence we have the set of equations

$$\frac{dP_0}{dx} = \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}x^2} P_1(x), \quad (3.12)$$

$$\frac{d^2P_0}{dx^2} + x \frac{dP_0}{dx} = \frac{1}{(2\pi)} e^{-x^2} P_2(x), \quad (3.13)$$

$$\frac{d^3P_0}{dx^3} + 3x \frac{d^2P_0}{dx^2} + (2x^2 + 1) \frac{dP_0}{dx} = \frac{1}{(2\pi)^{\frac{3}{2}}} e^{-\frac{3}{2}x^2} P_3(x), \quad (3.14)$$

$$\frac{d^4P_0}{dx^4} + 6x \frac{d^3P_0}{dx^3} + (11x^2 + 4) \frac{d^2P_0}{dx^2} + (6x^3 + 7x) \frac{dP_0}{dx} = \frac{1}{(2\pi)^2} e^{-2x^2} P_4(x), \quad (3.15)$$

$$\frac{d^5P_0}{dx^5} + 10x \frac{d^4P_0}{dx^4} + (35x^2 + 10) \frac{d^3P_0}{dx^3} + (50x^3 + 45x) \frac{d^2P_0}{dx^2} + (24x^4 + 46x^2 + 7) \frac{dP_0}{dx} = \frac{1}{(2\pi)^{\frac{5}{2}}} e^{-\frac{5}{2}x^2} P_5(x). \quad (3.16)$$

We can proceed in this manner up to any order, but it should be noted that $P_n(x)$ is a constant and $P_t(x) = 0$ if $t > n$. The general equation is

$$g_{0,t}(x) \frac{d^tP_0}{dx^t} + g_{1,t}(x) \frac{d^{t-1}P_0}{dx^{t-1}} + \dots + g_{t-1,t}(x) \frac{dP_0}{dx} = \frac{1}{(2\pi)^{\frac{1}{2}t}} e^{-\frac{1}{2}tx^2} P_t(x). \quad (3.17)$$

4. MOMENTS OF $X_{(k)}$

We shall first prove the following Lemma:

Lemma. If α and r are non-negative integers, then

$$\int_{-\infty}^{+\infty} x^\alpha \frac{d^{r+1}P_0}{dx^{r+1}} dx = (-1)^r \alpha(\alpha-1) \dots (\alpha-r+1) \mu'_{\alpha-r} \quad \text{or} \quad 0 \quad (4.1)$$

according as $r \leq \alpha$ or $r > \alpha$, where $\mu'_{\alpha-r}$ is the $(\alpha-r)$ th order moment of $X_{(k)}$ about the origin.

It should be noted that by definition

$$\int_{-\infty}^{+\infty} x^\alpha \frac{dP_0}{dx} dx = \mu'_\alpha. \quad (4.2)$$

$$\text{From (3.12) and (3.13)} \quad \frac{dP_0}{dx} = \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}x^2} P_1(x), \quad (4.3)$$

$$\frac{d^2P_0}{dx^2} = -\frac{x}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}x^2} P_1(x) + \frac{1}{2\pi} e^{-x^2} P_2(x), \quad (4.4)$$

and in general using the system of equations (3.12)–(3.17) we can write

$$\frac{d^rP_0}{dx^r} = \sum_{i=1}^r h_i(x) e^{-\frac{1}{2}ix^2} P_i(x), \quad (4.5)$$

where $h_i(x)$ is a polynomial in x . Now

$$\int_{-\infty}^{+\infty} x^\alpha \frac{d^{r+1}P_0}{dx^{r+1}} dx = \left| x^\alpha \frac{d^rP_0}{dx^r} \right|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} \alpha x^{\alpha-1} \frac{d^rP_0}{dx^r} dx. \quad (4.6)$$

Since $P_i(x)$ for any non-negative integer i is a bounded function of x , it follows from (4.5) that the first part on the right-hand side of (4.6) vanishes. Repeating this process we get if $r \leq \alpha$

$$\begin{aligned} \int_{-\infty}^{+\infty} x^\alpha \frac{d^{r+1}P_0}{dx^{r+1}} dx &= (-1)^r \alpha(\alpha-1) \dots (\alpha-r+1) \int_{-\infty}^{+\infty} x^{\alpha-r} \frac{dP_0}{dx} dx \\ &= (-1)^r \alpha(\alpha-1) \dots (\alpha-r+1) \mu'_{\alpha-r}. \end{aligned}$$

If $r > \alpha$, we get, on repeating the process α times,

$$\begin{aligned} \int_{-\infty}^{+\infty} x^\alpha \frac{d^{r+1}P_0}{dx^{r+1}} dx &= (-1)^\alpha \alpha(\alpha-1) \dots 3.2.1 \int_{-\infty}^{+\infty} \frac{d^{r+1-\alpha}P_0}{dx^{r+1-\alpha}} dx \\ &= (-1)^\alpha \alpha(\alpha-1) \dots 3.2.1 \left| \frac{d^{r-\alpha}P_0}{dx^{r-\alpha}} \right|_{-\infty}^{+\infty} = 0. \end{aligned}$$

This proves the Lemma.

On applying the Lemma and integrating the equations (3.13)–(3.16) we get

$$\mu'_1 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} P_2(x) e^{-x^2} dx, \quad (4.7)$$

$$-3 + (2\mu'_2 + 1) = \frac{1}{(2\pi)^{\frac{3}{2}}} \int_{-\infty}^{+\infty} P_3(x) e^{-\frac{3}{2}x^2} dx, \quad (4.8)$$

$$-22\mu'_1 + (6\mu'_3 + 7\mu'_1) = \frac{1}{(2\pi)^2} \int_{-\infty}^{+\infty} P_4(x) e^{-2x^2} dx, \quad (4.9)$$

$$70 - 150\mu'_2 - 45 + (24\mu'_4 + 46\mu'_2 + 7) = \frac{1}{(2\pi)^{\frac{5}{2}}} \int_{-\infty}^{+\infty} P_5(x) e^{-\frac{5}{2}x^2} dx. \quad (4.10)$$

We may write $\mu'_\alpha(n, k)$ instead of μ'_α to denote the fact that we have the α th moment about the origin of the k th order statistic out of a sample of n observations from $N(0, 1)$. We then have

$$\mu'_1(n, k) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} P_2(x) e^{-x^2} dx, \quad (4.11)$$

$$\mu'_2(n, k) = 1 + \frac{1}{2! (2\pi)^{\frac{3}{2}}} \int_{-\infty}^{+\infty} P_3(x) e^{-\frac{3}{2}x^2} dx, \quad (4.12)$$

$$\mu'_3(n, k) = \frac{5}{2} \mu'_1(n, k) + \frac{1}{3! (2\pi)^2} \int_{-\infty}^{+\infty} P_4(x) e^{-2x^2} dx, \quad (4.13)$$

$$\mu'_4(n, k) = -\frac{4}{3} + \frac{13}{3} \mu'_2(n, k) + \frac{1}{4! (2\pi)^{\frac{5}{2}}} \int_{-\infty}^{+\infty} P_5(x) e^{-\frac{5}{2}x^2} dx. \quad (4.14)$$

In general applying the Lemma to (3.17) we can express $\mu'_t(n, k)$ in terms of lower moments of even (odd) order when t is even (odd) and the integral

$$\int_{-\infty}^{+\infty} P_{t+1}(x) e^{-\frac{1}{2}(t+1)x^2} dx, \quad (4.15)$$

where the polynomials $P_2(x) \dots P_5(x)$ are given by (2.8)–(2.11). In the particular case when $n = k$, i.e., $x_{(n)}$ is the largest of $x_1, x_2, \dots, x_n$, $P_t(x)$ assumes the very simple form

$$P_t(x) = n(n-1)(n-2) \dots (n-t+1) [\Phi(x)]^{n-t}. \quad (4.16)$$

Hence we get

$$\mu'_1(n, n) = \frac{2 \binom{n}{2}}{2\pi} \int_{-\infty}^{+\infty} [\Phi(x)]^{n-2} e^{-x^2} dx, \quad (4.17)$$

$$\mu'_2(n, n) = 1 + \frac{3 \binom{n}{3}}{(2\pi)^{\frac{3}{2}}} \int_{-\infty}^{+\infty} [\Phi(x)]^{n-3} e^{-\frac{3}{2}x^2} dx, \quad (4.18)$$

$$\mu'_3(n, n) = \frac{5}{2} \mu'_1(n, n) + \frac{4 \binom{n}{4}}{(2\pi)^2} \int_{-\infty}^{+\infty} [\Phi(x)]^{n-4} e^{-2x^2} dx, \quad (4.19)$$

$$\mu'_4(n, n) = -\frac{4}{3} + \frac{13}{3} \mu'_2(n, n) + \frac{5 \binom{n}{5}}{(2\pi)^{\frac{5}{2}}} \int_{-\infty}^{+\infty} [\Phi(x)]^{n-5} e^{-\frac{5}{2}x^2} dx. \quad (4.20)$$

It should be noted that in the formulae (4.17) to (4.20) $[\Phi(x)]^{n-t}$ should be interpreted as zero if $t > n$.

Some integrals of the type occurring in this paper have been numerically evaluated by Hojo (1931). Equations (4.11), ..., (4.14), together with equations (2.8), ..., (2.11) of our paper correspond to equations (19), ..., (22) on pp. 334–5 of Hojo's paper. If we make the following correspondences between our and Hojo's paper:

$$k-1 \rightarrow p = \text{no. below } X_{(k)}, \quad n-k \rightarrow m = \text{no. above } X_{(k)},$$

our equation (4.11), together with (2.8), reduces to Hojo's equation (19) and so on.*

* [It appears that the functions T , I , S and K tabled by Hojo on pp. 325–6 will provide values of the 1st and 2nd moments of all order statistics of samples up to $n = 13$ inclusive and the 3rd and 4th moments up to $n = 10$ inclusive. His functions were, however, only calculated to 8 decimal places and when the raw moments are converted to moments about the mean, the effective number of

5. EXACT VALUES OF SOME MOMENTS

Let
$$I_n(a) = \int_{-\infty}^{+\infty} [\Phi(ax)]^n e^{-x^2} dx, \quad (5.1)$$

then
$$I_0(a) = \pi^{\frac{1}{2}}. \quad (5.2)$$

Now
$$\int_{-\infty}^{+\infty} [\Phi(ax) - \tfrac{1}{2}]^{2m+1} e^{-x^2} dx = 0, \quad (5.3)$$

since the integrand is an odd function of x . Hence

$$I_{2m+1}(a) = \frac{\sum_{r=1}^{2m+1} (-1)^{r+1} \binom{2m+1}{r} I_{2m-r+1}(a)}{2^r}. \quad (5.4)$$

In particular
$$I_1(a) = \tfrac{1}{2} I_0(a) = \tfrac{1}{2} \pi^{\frac{1}{2}}, \quad (5.5)$$

and
$$I_3(a) = \tfrac{3}{2} I_2(a) - \tfrac{3}{4} I_1(a) + \tfrac{1}{8} I_0(a) = \tfrac{3}{2} I_2(a) - \tfrac{1}{4} I_0(a). \quad (5.6)$$

In general, $I_{2m+1}(a)$ can be expressed as a linear function of $I_{2m}(a), I_{2m-2}(a), \dots, I_0(a)$.

Differentiating (5.1) with respect to a (this is justified in virtue of the uniform convergence of the integrals with respect to a , $-\infty < a < \infty$, and the continuity of the integrands), we get, for $n = 2$,

$$\frac{dI_2}{da} = \frac{1}{\pi^{\frac{1}{2}}} \frac{a}{(a^2 + 2)(a^2 + 1)^{\frac{1}{2}}}, \quad (5.7)$$

so that
$$I_2(a) = \frac{1}{\pi^{\frac{1}{2}}} \arctan(\sqrt{[a^2 + 1]}). \quad (5.8)$$

Using (5.6)
$$I_3(a) = \frac{3}{2\pi^{\frac{1}{2}}} \arctan(\sqrt{[a^2 + 1]}) - \tfrac{1}{4} \pi^{\frac{1}{2}}. \quad (5.9)$$

Since $P_{t+1}(x)$ is a polynomial in $\Phi(x)$ of degree $n - t - 1$, by using (5.2), (5.5), (5.8) and (5.9), we can exactly evaluate (4.15) if $n \leq t + 4$. Hence we can exactly evaluate $\mu'_t(n, k)$ for all odd values of t , if $n \leq 5$ and all even values of t if $n \leq 6$. Godwin (1949) and Jones (1948) have given tables of exact moments μ'_t for $t = 1$ and 2. The corresponding tables for $t = 3$ and 4 are given below.

Table 1. $\mu'_3(n, k)$

$k \backslash n$	$k = n$	$k = n - 1$	$k = n - 2$
2	$2A = 1.41047 \ 39589$	—	—
3	$3A = 2.11571 \ 09383$	0	—
4	$2B + 2C = 2.70042 \ 57044$	$12A - 6B - 6C = 0.36156 \ 66401$	—
5	$-5A + 5B + 5C = 3.22487 \ 93638$	$20A - 10B - 10C = 0.60261 \ 10668$	0

Here:
$$A = \frac{5}{4\sqrt{\pi}} = 0.70523 \ 69794 \ 35, \quad B = \frac{1}{2\sqrt{(2\pi)^{\frac{3}{2}}}} = 0.06349 \ 36359 \ 34,$$

$$C = \frac{15}{2\pi^{\frac{3}{2}}} \arctan(\sqrt{2}) = 1.28671 \ 92162 \ 55.$$

significant figures is very much reduced. As M. G. Kendall has pointed out (*Biometrika*, 1954, 41, p. 561) Hojo's values for his S -functions as printed appear to be in error by the omission of a factor of 2π which seems to have been included correctly in his moment calculations.—Ed.)

Table 2. $\mu'_4(n, k)$

$k \backslash n$	$k = n$	$k = n - 1$	$k = n - 2$
2	3.0	—	—
3	$3 + a$ = 4.19454 59401	$3 - 2a$ = 0.61090 81198	—
4	$3 + 2a$ = 5.38909 18802	$3 - 2a$ = 0.61090 81198	—
5	$3 + b + c$ = 6.52339 54861	$3 + 10a - 4b - 4c$ = 0.85187 74565	$3 - 20a + 6b + 6c$ = 0.24945 41149
6	$3 - 5a + 3b + 3c$ = 7.59745 67578	$3 + 25a - 9b - 9c$ = 1.15308 91273	$3 - 20a + 6b + 6c$ = 0.24945 41149

Here: $a = \frac{13}{\sqrt{3}(2\pi)} = 1.19454\ 59400\ 81$, $b = \frac{65}{\sqrt{3}\pi^2} \arctan(\sqrt{\frac{5}{3}}) = 3.46675\ 52225\ 38$,
 $c = \frac{\sqrt{5}}{4\pi^2} = 0.05664\ 02635\ 46$.

The authors wish to thank Miss Phyllis A. Groll of Bell Telephone Laboratories for numerical computation in Tables 1 and 2.

REFERENCES

- BAHADUR, RAGHU RAJ (1950). On a problem in the theory of k populations. *Ann. Math. Statist.* **21**, 362-75.
- BECHHOFFER, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.* **25**, 16-29.
- BECHHOFFER, R. E., DUNNETT, C. W. & SOBEL, M. (1954). A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance. *Biometrika*, **41**, 170-6.
- BECHHOFFER, R. E. & SOBEL, M. (1956). A sequential multiple decision procedure for selecting the population with the largest mean from k normal populations with a common unknown variance (preliminary report). Abstract, *Ann. Math. Statist.* **27**, 218-19.
- COLE, R. H. (1951). Relations between moments of order statistics. *Ann. Math. Statist.* **20**, 279-85.
- GODWIN, H. J. (1949). Some low moments of order statistics. *Ann. Math. Statist.* **20**, 279-85.
- GODWIN, H. J. (1949). On the estimation of dispersion by linear systematic statistics. *Biometrika*, **36**, 92-100.
- GUPTA, SHANTI S. (1956). On a decision rule for a problem in ranking means. Ph.D. Thesis, University of North Carolina, Chapel Hill.
- HASTINGS, C. JR., MOSTELLER, F., TUKEY, J. W. & WINSOR, C. P. (1947). Low moments for small samples: a comparative study of order statistics. *Ann. Math. Statist.* **18**, 413-26.
- HOJO, T. (1931). Distribution of median, quartiles, and interquartile distance in samples from a normal population. *Biometrika*, **23**, 315-60.
- JONES, H. L. (1948). Exact lower moments of order statistics in small samples from a normal distribution. *Ann. Math. Statist.* **19**, 270-3.
- MOSTELLER, F. (1946). On some useful inefficient statistics. *Ann. Math. Statist.* **17**, 377-408.
- MOSTELLER, F. (1948). A k -sample slippage test for an extreme population. *Ann. Math. Statist.* **19**, 58-65.

the shortest unbiased system of confidence intervals (Eudey, 1949; Blank, 1956). Stevens (1957) has given a practical procedure for calculating such a randomized interval for either the binomial or Poisson parameters. Although his procedure is short (of the order of 15 min. in a trial by the present writers), the almost universal occurrence of the Poisson distribution seems to warrant the publication of a table giving an optimum interval directly.

2. A PROPERTY OF POISSON PROBABILITY SUMS

The Poisson distribution is the distribution of a random variable c which has the expectation m and the probability

$$p_i(m) = \frac{e^{-m} m^i}{i!} \quad (i = 0, 1, 2, \dots; m > 0) \quad (1)$$

of taking on the value i . We note the well-known fact that for fixed m the sequence (1) has a unique largest number except that when m is an integer the maximum is attained by two consecutive members. On each side of the largest member(s) $p_i(m)$ is a strictly monotone function of i .

On the other hand, a sum

$$\sum_{i=c_1}^{c_2} p_i(m) \quad (c_1 \leq c_2)$$

over a fixed set $(c_1, c_1 + 1, \dots, c_2)$ is an analytic function of m and has a unique maximum for m satisfying

$$\frac{d}{dm} \sum_{i=c_1}^{c_2} p_i(m) \equiv p_{c_1-1}(m) - p_{c_2}(m) = 0 \quad (c_1 > 0), \quad (2)$$

that is, for

$$m = m_{c_1, c_2} \equiv [c_1(c_1 + 1) \dots c_2]^{1/(c_2 - c_1 + 1)}. \quad (3)$$

(Equation (3) applies for $c_1 = 0$, as does (2) if $p_{-1}(m)$ is defined to be identically zero.) Since $p_{c_1-1}(m) \cong p_{c_2}(m)$ according as $m \leq m_{c_1, c_2}$, ($c_1 > 0$), it follows that

$$\sum_{i=c_1-1}^{c_2-1} p_i(m) \cong \sum_{i=c_1}^{c_2} p_i(m) \quad \text{according as} \quad m \leq m_{c_1, c_2}. \quad (4)$$

We note that m_{c_1, c_2} increases as either one of its subscripts increases. From (4) we see that, for m between m_{c_1-1, c_2-1} and m_{c_1, c_2} ,

$$\sum_{i=c_1-1}^{c_2-1} p_i(m) > \sum_{i=c_1}^{c_2} p_i(m) > \sum_{i=c_1+1}^{c_2+1} p_i(m) > \dots$$

Also for this interval

$$\sum_{i=c_1-1}^{c_2-1} p_i(m) \geq \sum_{i=c_1-2}^{c_2-2} p_i(m) > \sum_{i=c_1-3}^{c_2-3} p_i(m) > \dots$$

Hence for fixed r

$$\text{Max}_{c=0, 1, \dots} \sum_{i=c}^{c+r} p_i(m) = \sum_{i=c_1}^{c_1+r} p_i(m) \quad (m_{c_1, c_1+r} \leq m \leq m_{c_1+1, c_1+r+1}; c_1 = 0, 1, \dots; r = 0, 1, \dots), \quad (5)$$

and these maximum sums form a non-increasing function of m , as illustrated in Fig. 1.

3. CALCULATION OF THE TABLE

The calculation of Table 1 printed on pp. 448-53 below is of the same general nature as that described in detail by Crow (1956) for the binomial parameter. The main body of calculation was performed on a high-speed electronic computer which calculated in the process all required terms of the Poisson distribution from (1), without reference to existing tables. The confidence limits for $c = 0(1)20$ were checked by independent calculation using a desk

calculator and the tables of Poisson individual terms and sums of Molina (1942) and Pearson & Hartley (1954), and Kitagawa's (1952) table of individual terms. Both calculations were made to three decimal places of the parameter m , but accuracy of the last place was not always sought in the desk calculation.

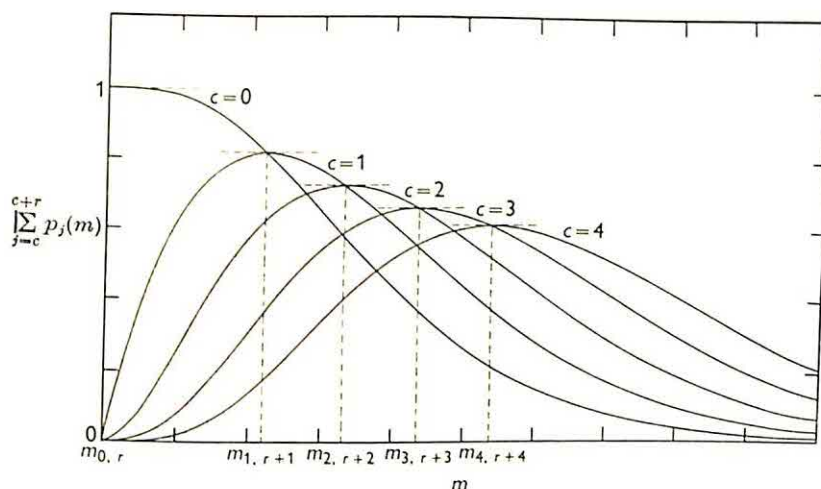


Fig. 1. Poisson probability sums illustrated with $r = 3$. The characteristics are general except that for $r = 0$, $c = 0$ the slope is negative at $m = 0$.

The calculation of a system, say $\delta \equiv \delta(c; m, \epsilon)$, of confidence intervals for m with confidence coefficient $1 - \epsilon$ from observations on the Poisson variable c reduces to the determination, for each m , of acceptance regions $A(m)$ consisting of consecutive integers $c_1 \geq 0, c_1 + 1, \dots, c_2$ such that

$$\sum_{i=c_1}^{c_2} p_i(m) \geq 1 - \epsilon \quad (6)$$

and satisfying alternative further conditions sufficient to define the system uniquely. The central confidence intervals previously tabulated will be denoted by δ_1 and the system in Table 1, defined by the two properties stated in § 1, will be denoted by δ_4 (following Crow's notation for the analogous binomial intervals). The system δ_3 is defined by Sterne's (1954) condition that each $p_i(m)$ included in the sum (6) be not smaller than each $p_i(m)$ excluded; an immediate consequence of this condition is that the corresponding acceptance regions $A_3(m)$ are as short as possible, in other words, that δ_3 has property (i) stated in § 1. The system δ_4 arose as a modification of δ_3 which is made possible by the discreteness of c and the continuity of the sum (6) as a function of m for fixed c_1 and c_2 .

The acceptance regions $A_3(m)$ and $A_4(m)$ are determined by proceeding continuously from small to larger values of m , starting at $m = 0$, and making a change in c_1 or c_2 or both in (6) as required. If for a given m , say m' , the acceptance region $A_i(m)$, ($i = 3, 4$), is $(c'_1, \dots, c'_2 = c'_1 + r')$, then by definition of the $A_i(m)$ no probability sum of less than $r' + 1$ terms is as large as $1 - \epsilon$ for either $m = m'$ or, consequently, by equation (5) and the concluding statement of § 2, for $m > m'$. Hence the size of the acceptance region $A_i(m)$ ($i = 3, 4$) cannot decrease as m increases. (This is not true for the corresponding binomial distribution acceptance regions.) We therefore maintain the same size if possible by substituting $c'_2 + 1$ for c'_1 , at $m = m_{c'_1+1, c'_2+1}$ for $A_3(m)$, or where $\sum_{i=c'_1+1}^{c'_2+1} p_i(m)$ first attains $1 - \epsilon$ for $A_4(m)$. If the probability sum drops to $1 - \epsilon$ before substitution is possible, we enlarge $A_i(m)$ by adding

$c'_2 + 1$ to $(c'_1, \dots, c'_2)$. Alternatively, it would be possible to shift or enlarge the acceptance region to the left, adding $c'_1 - 1$, but to no useful purpose; in fact it would yield confidence sets that are pairs of intervals rather than single intervals.*

4. COMPARISON OF SYSTEMS OF CONFIDENCE INTERVALS

For the purpose of comparison the δ_3 system was also computed for $c \leq 50$ and the δ_1 system obtained from Pearson & Hartley's tables 40 and 8. Some lengths d_1 of δ_1 and corresponding algebraic and percentage decreases in length from d_1 are shown in Table 2 for confidence coefficients of 90 and 99 %. The results are rounded.

Table 2. Comparison of d_1 , d_3 and d_4

c	d_1		$d_1 - d_4$		$100(d_1 - d_4)/d_1$		$100(d_1 - d_3)/d_1$	
	90	99	90	99	90	99	90	99
0	3.00	5.30						
1	4.69	7.43	0.56	0.53	18.7 %	9.9 %	0.1 %	0.1 %
2	5.94	9.17	0.27	0.52	5.7	7.0	-3.3	1.3
3	6.94	10.64	0.50	0.59	8.4	6.5	0.0	0.0
4	7.79	11.92	0.53	0.60	7.6	5.7	0.9	2.0
5	8.54	13.07	0.94	0.37	12.0	3.1	6.4	-0.5
6	9.23	14.12	1.26	0.56	14.7	4.3	5.4	0.3
7	9.86	15.10	0.32	0.63	3.5	4.5	2.4	0.7
8	10.45	16.01	0.92	0.63	9.3	4.1	4.2	0.7
9	11.01	16.87	1.43	0.55	13.7	3.4	4.5	0.7
10	11.54	17.68	0.24	0.91	2.2	5.4	3.8	2.9
12	12.52	19.20	1.52	1.13	13.2	6.4	4.3	2.5
14	13.42	20.61	1.63	0.21	13.0	1.1	4.1	0.9
16	14.26	21.91	0.99	1.28	7.4	6.2	3.3	2.0
18	15.06	23.15	0.15	0.81	1.1	3.7	1.9	1.3
20	15.81	24.32	0.24	0.04	1.6	0.2	3.4	0.6
25	17.54	27.00	1.27	0.61	8.0	2.5	2.1	1.4
30	19.10	29.44	1.71	0.42	9.7	1.6	3.2	1.6
40	21.87	33.77	0.03	1.30	0.2	4.4	2.9	1.5
50	24.33	37.61	0.90	1.25	4.1	3.7	1.6	0.7
			0.08	1.74	0.3	4.6	1.4	1.1

The differences $d_1 - d_4$ (observed for confidence coefficient 95 % as well as those listed above and for somewhat more values of c than listed) oscillate rapidly within the range 0 to 2 as c varies, with no substantial trend as either c or confidence coefficient varies. Consequently the relative improvement of δ_4 over δ_1 is appreciable for at least small values of c but tends to decrease as c increases. The relative improvement of δ_4 over δ_3 is also appreciable for small values of c but occasionally becomes negative. However, since δ_3 and δ_4 have acceptance regions of the same size, the average of $(d_1 - d_4)/d_1$ over $c = 0, 1, \dots$ to any finite value exceeds the corresponding average of $(d_1 - d_3)/d_1$.

* The confidence sets for the parameter of a binomial distribution considered by Clunies-Ross (1958) occasionally consist of pairs of intervals, including cases for confidence coefficients 0.90 and 0.99. Hence his statement on p. 278: 'In practice it seems that the restriction (4.6 a, b) is not needed for the usual values of $\alpha = 0.9500, 0.9000$ ', needs correction; see Crow (1956), p. 425.

It is of interest to compare the tabulated δ_4 limits with various large-sample approximations in order to test the validity of using the δ_4 limits as far as they are tabulated and an approximation beyond that. Such a comparison is summarized below for confidence coefficient 95 % and $c = 50$:

Table 3. *Comparison with large-sample approximations*

Identification of interval	Eq. no. in Hald (1952)	Upper limit	Lower limit	Length of interval	% Diff. of length from d_1
δ_1	—	65.92	37.11	28.81	0.0
δ_3	—	65.85	37.67	28.18	-2.2
δ_4	—	64.95	37.67	27.28	-5.3
Normal approx. without continuity correction	—	65.91	37.93	27.98	-2.9
Normal approx. with continuity correction	(22.5.4)	66.48	37.50	28.98	+0.6
Square-root normal approx. with continuity correction	(22.5.5)	65.39	36.67	28.72	-0.3
Mean of (22.5.4) and (22.5.5)	—	65.93	37.08	28.85	+0.1

Quite generally the continuity correction lengthens the interval by very close to 1. The mean of the two approximations with continuity corrections is closer to δ_1 than either one singly for both upper and lower limits in the above instance, and this seems to be the case for $c > 50$ and even substantially smaller c (cf. Hald, p. 718). Consequently the suggested large-sample approximation, effectively identical with the mean of (22.5.4) and (22.5.5), but simpler by virtue of averaging the two approximations to percentage points for a given m (Hald's (22.2.7)) before solving for the limits, is

$$\left. \begin{matrix} m_U \\ m_L \end{matrix} \right\} \cong c \pm \frac{1}{2} + \frac{3}{8}u_\epsilon^2 \pm u_\epsilon \sqrt{(c \pm \frac{1}{2} + \frac{1}{8}u_\epsilon^2)}, \quad (7)$$

where u_ϵ is the upper $100\epsilon/2$ percentage point of the normal distribution of mean 0 and variance 1, and the upper signs are for the upper limit m_U , the lower for the lower limit m_L . Equation (7) is a remarkably accurate approximation to δ_1 ; the absolute error in m_U or m_L apparently is less than 0.1 for $c > 0$ when $\epsilon = 0.10$ and for $c > 2$ when $\epsilon = 0.05$. An alternative approach is that of Walsh (1954), who considered, rather than the correction terms of (7), functions of ϵ which he then determined so as to render probability errors small over a wide range. Comparisons similar to the above for confidence coefficients of 90 and 99 % show about the same order of differences. Since the δ_4 intervals are shorter than δ_1 intervals in all observed cases and the percentage difference in length tends to zero as c becomes infinite, it is concluded that using δ_4 intervals as far as they are tabulated and the large-sample approximation (7) for larger c is a valid and satisfactory procedure.

For comparison with δ_4 the randomized 90 % confidence interval was calculated according to the procedure of Stevens (1957) for the first few values of c and intervals of 0.1 in the added

uniformly distributed variable.* The mean lengths, say $\bar{d}_5$, of these intervals for each c are compared in Table 4 with the corresponding values of d_4 :

Table 4. Comparison with mean randomized intervals

c	$\bar{d}_5$	d_4	$100(\bar{d}_5 - d_4)/d_4$
0	2.04	2.44	-16.4%
1	3.90	4.43	-12.0
2	5.11	5.44	-6.1
3	6.08	6.41	-5.1
4	6.91	6.85	+0.9
5	7.66	7.28	+5.2
6	8.33	8.91	+6.5
7	8.96	8.94	+0.2
8	9.54	9.02	+5.8
9	10.10	10.77	-6.2
10	10.62	10.01	+6.1

The system δ_4 varies irregularly in length and location while $\bar{d}_5$ varies smoothly. However, in any particular case δ_5 is determined in part by a random number x between 0 and 1 and hence will take on some irregularity; as an extreme example, the 90% interval for $c = 0$ and $x \leq 0.05$ has length 0.

Crow (1956) discussed briefly, in relation to the binomial parameter, the use of the expected length of a confidence interval as a criterion for choice of interval. In a similar

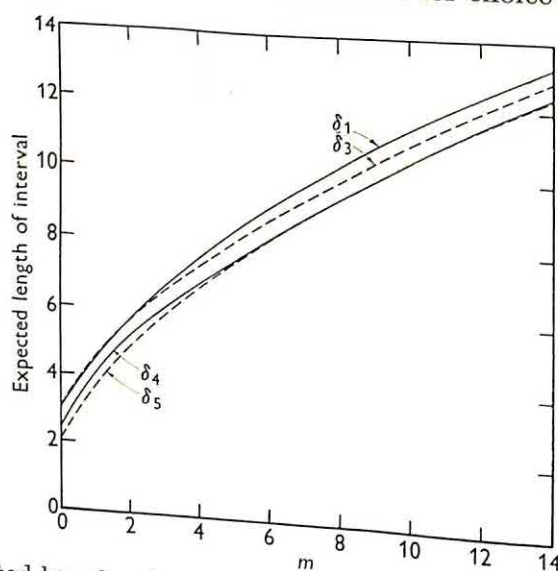


Fig. 2. Expected lengths of various confidence intervals for 90% confidence. δ_1 : central; δ_3 : Sterne-type; δ_4 : tabulated herein; δ_5 : randomized central.

* A slight change from his procedure was in using for $c = 0, 1, 2, 3$, the exact formula (Stevens, 1950) rather than the quadratic Taylor's series approximation. The approximation is not always accurate to within 0.1 for such small c , so that it is desirable to calculate both the forward extrapolate (from $x = 0$, in Stevens's notation) and the backward extrapolate (from $x = 1$). The true limit is usually between the two extrapolates, but one exception was found. The forward extrapolate is more accurate than the backward extrapolate.

manner the expected lengths of δ_1 , δ_3 , δ_4 , and the shortest unbiased system, say δ_5 , were calculated for small values of m , the last one by Stevens's procedure, using Molina's table; these expected lengths are graphed in Fig. 2. The randomized system δ_5 has the smallest expected length in the range of calculation, but the difference from that of δ_4 is not more than 0.06 for $6 \leq m \leq 14$. The expected length of δ_3 lies between those of δ_1 and δ_4 except that it exceeds that of δ_1 slightly for $0.01 < m < 1.5$. By using δ_4 instead of δ_1 the experimenter can 'expect' a 90% confidence interval shorter by between 0.45 (for $m = 1$) and about 0.9 (for $7 \leq m \leq 14$).

REFERENCES

- BLANK, A. A. (1956). Existence and uniqueness of a uniformly most powerful randomized unbiased test for the binomial. *Biometrika*, **43**, 465-6.
- CLUNIES-ROSS, C. W. (1958). Interval estimation for the parameter of a binomial distribution. *Biometrika*, **45**, 275-9.
- CROW, E. L. (1956). Confidence intervals for a proportion. *Biometrika*, **43**, 423-35.
- EUDEY, M. W. (1949). On the treatment of discontinuous random variables. Technical Report No. 13. Statistical Laboratory, University of California, Berkeley.
- GARWOOD, F. (1936). Fiducial limits for the Poisson distribution. *Biometrika*, **28**, 437-42.
- HALD, A. (1952). *Statistical Theory with Engineering Applications*. New York: John Wiley and Sons Inc.
- KITAGAWA, T. (1952). *Tables of the Poisson Distribution*. Tokyo: Baifukan.
- MOLINA, E. C. (1942). *Poisson's Exponential Binomial Limit*. New York: D. van Nostrand Co. Inc.
- PEARSON, E. S. & HARTLEY, H. O. (1954). *Biometrika Tables for Statisticians*, **1**. Cambridge University Press.
- RICKER, W. E. (1937). The concept of confidence or fiducial limits applied to the Poisson frequency distribution. *J. Amer. Statist. Ass.* **32**, 349-56.
- STERNE, T. E. (1954). Some remarks on confidence or fiducial limits. *Biometrika*, **41**, 275-8.
- STEVENS, W. L. (1950). Fiducial limits of the parameter of a discontinuous distribution. *Biometrika*, **37**, 117-39.
- STEVENS, W. L. (1957). Shorter intervals for the parameter of the binomial and Poisson distributions. *Biometrika*, **44**, 436-40.
- WALSH, J. E. (1954). Analytic tests and confidence intervals for the mean value, probabilities, and percentage points of a Poisson distribution. *Sankhyā*, **14**, 25-38.

Table 1. Confidence limits for the expectation of a Poisson variable

Confidence coefficient, $100(1 - \epsilon) \%$

c	80		90		95		99		99.9	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
0	0	1.819	0	2.436	0	3.285	0	4.771	0	7.065
1	0.223	3.546	0.105	4.532	0.051	5.323	0.010	6.914	0.001	9.561
2	0.824	4.758	0.532	5.976	0.355	6.686	0.149	8.727	0.045	11.532
3	1.535	5.882	1.102	7.512	0.818	8.102	0.436	10.473	0.190	13.596
4	1.819	7.564	1.745	8.597	1.366	9.598	0.823	12.347	0.429	15.144
5	2.645	8.529	2.433	9.716	1.970	11.177	1.279	13.793	0.739	17.114
6	3.546	9.922	2.436	11.342	2.613	12.817	1.785	15.277	1.107	18.490
7	3.914	10.969	3.589	12.531	3.285	13.765	2.330	16.801	1.520	19.987
8	4.758	12.481	4.532	13.553	3.285	14.921	2.906	18.362	1.971	21.603
9	5.696	13.243	4.532	15.298	4.460	16.768	3.507	19.462	2.452	23.170
10	5.882	15.205	5.976	15.985	5.323	17.633	4.130	20.676	2.961	24.677
11	7.564	15.438	5.976	17.810	5.323	19.050	4.771	22.042	3.491	26.126
12	7.564	16.914	7.512	18.403	6.686	20.335	4.771	23.765	4.042	27.530
13	8.529	18.537	7.512	20.054	6.686	21.364	5.829	24.925	4.611	28.907
14	9.922	18.938	8.597	21.035	8.102	22.945	6.668	25.992	5.195	30.372
15	9.922	20.414	9.484	22.258	8.102	23.762	6.914	27.718	5.794	31.993
16	10.969	22.037	9.716	23.824	9.598	25.400	7.756	28.852	6.405	33.622
17	12.481	22.326	11.342	24.452	9.598	26.306	8.727	29.900	7.028	34.745
18	12.481	23.744	11.342	26.158	11.177	27.735	8.727	31.839	7.065	35.916
19	13.243	25.707	12.531	26.935	11.177	28.966	10.009	32.547	8.023	37.384
20	15.205	25.707	13.553	28.092	12.817	30.017	10.473	34.183	8.840	39.108
21	15.205	26.972	13.553	29.988	12.817	31.675	11.242	35.204	9.561	40.124
22	15.438	28.469	15.298	30.179	13.765	32.277	12.347	36.544	9.561	41.245
23	16.914	29.983	15.795	31.639	14.921	34.048	12.347	37.819	10.710	43.041
24	18.537	30.152	15.985	33.444	14.921	34.665	13.793	38.939	11.532	44.162
25	18.537	31.507	17.810	33.643	16.768	36.030	13.793	40.373	11.532	45.213
26	18.94	33.03	18.28	35.08	16.77	37.67	15.28	41.39	12.73	47.08
27	20.41	34.42	18.40	37.00	17.63	38.16 ⁵	15.28	42.85	13.60	48.01
28	22.04	34.58 ⁵	20.05	37.04	19.05	39.76	16.80	43.91	13.60	49.32
29	22.04	35.92	21.03 ⁵	38.44	19.05	40.94	16.80	45.26	14.89	50.97
30	22.33	37.39	21.03 ⁵	40.10 ⁵	20.33 ⁵	41.75	18.36	46.50	15.14	51.76
31	23.74	39.07	22.26	40.99	21.36	43.45	18.36	47.62	16.11	53.54
32	25.71	39.07	23.82	41.74	21.36	44.26	19.46	49.13	17.11	54.45 ⁵
33	25.71	40.23 ⁵	23.82	43.22	22.94 ⁵	45.28	20.28 ⁵	49.96	17.11	55.88
34	25.71	41.62	24.45	44.87	23.76	47.02 ⁵	20.68	51.78	18.49	57.15
35	26.97	43.25	26.16	45.00	23.76	47.69	22.04	52.28	18.49	58.20
36	28.47	44.20	26.93 ⁵	46.38	25.40	48.74	22.04	54.03	19.87	59.87 ⁵
37	29.98	44.48	26.93 ⁵	47.97	26.31	50.42	23.76 ⁵	54.74	19.99	60.63
38	29.98	45.79	28.09	49.12	26.31	51.29	23.76 ⁵	56.14	21.27	62.67
39	30.15	47.20	29.99	49.56	27.73 ⁵	52.15	24.92 ⁵	57.61 ⁵	21.60	63.18
40	31.51	48.99	29.99	50.96	28.97	53.72	25.83	58.35	22.68	65.13
41	33.03	49.46	30.18	52.64	28.97	54.99	25.99	60.39	23.17	65.70
42	34.42	49.94 ⁵	31.64	53.46	30.02	55.51	27.72	60.59	24.13	67.62
43	34.42	51.25	33.44	54.05	31.67 ⁵	56.99	27.72	62.13	24.68	68.19
44	34.58 ⁵	52.64	33.44	55.44 ⁵	31.67 ⁵	58.72	28.85	63.63 ⁵	25.63	70.12 ⁵
45	35.92	54.29	33.64	57.10	32.28	58.84	29.90	64.26	26.13	70.66
46	37.39	55.16	35.08	57.99	34.05	60.24	29.90	65.96	27.18	72.66
47	39.07	55.33	37.00	58.48	34.66 ⁵	61.90	31.84	66.81 ⁵	27.53	73.10
48	39.07	56.61	37.00	59.85	34.66 ⁵	62.81	31.84	67.92	28.76	74.96
49	39.07	57.95	37.04	61.41	36.03	63.49	32.55	69.83	28.91	75.52
50	40.23 ⁵	59.44	38.44	62.69	37.67	64.95	34.18	70.05	30.37	77.22

If c is the observed frequency or count and m_L, m_U are the lower and upper $100(1 - \epsilon) \%$ confidence limits for its expectation, m , then $\Pr(m_L \leq m \leq m_U) \geq 1 - \epsilon$.

Confidence coefficient, $100(1-\epsilon)\%$

c	80		90		95		99		99.9	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
51	41.62	61.02	40.10 ⁵	62.86	37.67	66.76	34.18	71.56	30.37	78.14
52	43.25	61.02	40.99	64.19	38.16 ⁵	66.76	35.20	73.20	31.99	79.53
53	44.20	61.90	40.99	65.64 ⁵	39.76	68.10	36.54	73.62	31.99	80.83
54	44.20	63.19	41.74	67.45	40.94	69.62	36.54	75.16	33.62	81.84 ⁵
55	44.48	64.56	43.22	67.45	40.94	71.09	37.82	76.61	33.62	83.50
56	45.79	66.12	44.87	68.49	41.75	71.28	38.94	77.15	34.74 ⁵	84.17
57	47.20	67.33	44.87	69.86	43.45	72.66	38.94	78.71	35.59	86.16
58	48.99	67.33	45.00	71.42	44.26	74.22	40.37	80.06	35.92	86.49
59	49.46	68.38	46.38	72.67	44.26	75.49	41.39	80.65	37.38	88.14
60	49.46	69.67 ⁵	47.97	72.76	45.28	75.78 ⁵	41.39	82.21	37.38	89.23
61	49.94 ⁵	71.04	49.12	74.07	47.02 ⁵	77.16	42.85	83.56	29.11	90.34
62	51.25	72.58 ⁵	49.12	75.47	47.69	78.73	43.91	84.12	39.11	92.01
63	52.64	73.87	49.56	77.15 ⁵	47.69	79.98	43.91	85.65	40.12	92.59
64	54.29	73.87	50.96	77.94	48.74	80.25	45.26	87.12	41.21	94.36
65	55.16	74.79	52.64	78.27	50.42	81.61	46.50	87.55	41.24 ⁵	95.01 ⁵
66	55.16	76.07	53.46	79.59	51.29	83.14	46.50	89.05	43.04	96.40
67	55.33	77.40	53.46	80.99	51.29	84.57	47.62	90.72	43.04	97.95
68	56.61	78.85	54.05	82.72	52.15	84.67	49.13	90.96	44.16	98.59
69	57.95	80.60	55.44 ⁵	83.37	53.72	86.01	49.13	92.42	45.18	100.36
70	59.44	80.60	57.10	83.73	54.99	87.48	49.96	94.34 ⁵	45.21	101.02
71	61.02	81.13	57.99	85.04	54.99	89.23	51.78	94.35	47.08	102.34
72	61.02	82.38	57.99	86.43	55.51	89.23	51.78	95.76	47.08	104.01
73	61.02	83.67	58.48	88.05	56.99	90.37	52.28	97.42	48.01	104.50
74	61.90	85.04	59.85	89.04	58.72	91.78	54.03	98.36	49.32	106.13
75	63.19	86.58	61.41	89.15	58.72	93.48	54.74	99.09	49.32	107.23 ⁵
76	64.56	87.87	62.69	90.44	58.84	94.23	54.74	100.61	50.97	108.17
77	66.12	87.87	62.69	91.79	60.24	94.70 ⁵	56.14	102.16 ⁵	51.39	110.17
78	67.33	88.64	62.86	93.29	61.90	96.06	57.61 ⁵	102.42	51.76	110.34
79	67.33	89.89	64.19	94.80	62.81	97.54 ⁵	57.61 ⁵	103.84	53.54	111.85
80	67.33	91.19	65.64 ⁵	94.80	62.81	99.17	58.35	105.66	53.54	113.51
81	68.38	92.55	67.45	95.78	63.49	99.17	60.39	106.12	54.45 ⁵	113.95
82	69.67 ⁵	94.11 ⁵	67.45	97.10	64.95	100.32	60.39	107.10	55.88	115.52
83	71.04	95.34	67.45	98.51	66.76	101.71	60.59	108.61 ⁵	55.88	116.84
84	72.58 ⁵	95.34	68.49	100.28	66.76	103.31 ⁵	62.13	110.16	57.15	117.54
85	73.87	96.08	69.86	100.81	66.76	104.40	63.63 ⁵	110.37	58.20	119.21
86	73.87	97.32	71.42	101.09 ⁵	68.10	104.58	63.63 ⁵	111.78	58.20	120.17
87	73.87	98.60	72.67	102.38	69.62	105.90 ⁵	64.26	113.45	59.87 ⁵	121.12
88	74.79	99.95	72.67	103.71	71.09	107.32	65.96	114.33	60.49	122.92
89	76.07	101.43	72.76	105.17	71.09	109.11	66.81 ⁵	114.99	60.63	123.46
90	77.40	103.02	74.07	106.92	71.28	109.61	66.81 ⁵	116.44	62.67	124.68
91	78.85	103.02	75.47	106.92	72.66	110.11	67.92	118.33	62.67	126.67
92	80.60	103.45	77.15 ⁵	107.63	74.22	111.44	69.83	118.33	63.18	126.75
93	80.60	104.68	77.94	108.91 ⁵	75.49	112.87	69.83	119.59	65.13	128.21
94	80.60	105.94	77.94	110.26	75.49	114.84	70.05	121.09	65.13	130.14
95	81.13	107.24	78.27	111.75	75.78 ⁵	114.84	71.56	122.69	65.70	130.26
96	82.38	108.63	79.59	113.35	77.16	115.60 ⁵	73.20	122.78	67.62	131.72
97	83.67	110.27 ⁵	80.99	113.35	78.73	116.93	73.20	124.16	67.62	133.64
98	85.04	111.18	82.72	114.11	79.98	118.35	73.62	125.70	68.19	133.75
99	86.58	111.18	83.37	115.40	79.98	120.36	75.16	127.07	70.12 ⁵	135.20
100	87.87	111.98	83.37	116.74	80.25	120.36	76.61	127.31	70.12 ⁵	137.15

Table 1 (cont.)

Confidence coefficient, $100(1-\epsilon)\%$

c	80		90		95		99		99.9	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
101	87.88	113.21	83.73	118.21	81.61	121.06	76.61	128.70	70.66	137.22
102	87.88	114.47 ⁵	85.04	119.88	83.14	122.37	77.15	130.27 ⁵	72.66	138.66
103	88.64	115.79	86.43	119.88	84.57	123.77	78.71	131.50	72.66	140.64
104	89.89	117.19	88.05	120.55	84.57	125.46	80.06	131.82	73.10	140.74
105	91.19	118.95	89.04	121.82	84.67	126.26	80.06	133.21	74.96	142.10
106	92.55	119.50	89.04	123.15	86.01	126.48	80.65	134.79	75.41	143.83
107	94.11 ⁵	119.50	89.15	124.57	87.48	127.78	82.21	135.99	75.52	144.55
108	95.34	120.44	90.44	126.52	89.23	129.14	83.56	136.30	77.22	145.51 ⁵
109	95.34	121.67	91.79	126.52	89.23	130.68	83.56	137.68	78.14	147.13 ⁵
110	95.34	122.92 ⁵	93.29	126.94 ⁵	89.23	132.03	84.12	139.24	78.14	148.24
111	96.08	124.23	94.80	128.20	90.37	132.03	85.65	140.54	79.53	148.92
112	97.32	125.61	94.80	129.50	91.78	133.14 ⁵	87.12	140.76	80.83	150.46
113	98.60	127.22	94.80	130.87	93.48	134.48	87.12	142.12	80.83	151.91
114	99.95	128.23	95.78	132.44	94.23	135.92	87.55	143.64	81.84 ⁵	152.31
115	101.43	128.23	97.10	133.66	94.23	137.79	89.05	145.13	83.50	153.79
116	103.02	128.84	98.51	133.66	94.70 ⁵	137.79	90.72	145.19	83.50	155.57 ⁵
117	103.02	130.06	100.28	134.54	96.06	138.49	90.72	146.54	84.17	155.69
118	103.02	131.30	100.81	135.81	97.54 ⁵	139.79	90.96	148.01	86.16	157.12
119	103.45	132.57 ⁵	100.81	137.13	99.17	141.16	92.42	149.76	86.16	158.89
120	104.68	133.91	101.09 ⁵	138.55	99.17	142.70	94.34 ⁵	149.76	86.49	159.49
121	105.94	135.37	102.38	140.54	99.17	144.01	94.34 ⁵	150.93	88.14	160.45
122	107.24	137.08	103.71	140.54	100.32	144.01	94.35	152.35 ⁵	89.23	162.01
123	108.63	137.08	105.17	140.85	101.71	145.08	95.76	154.18	89.23	163.36
124	110.27 ⁵	137.18	106.92	142.09	103.31 ⁵	146.39	97.42	154.60	90.34	163.79
125	111.18	138.38	106.92	143.37	104.40	147.80	98.36	155.31	92.01	165.25
126	111.18	139.60 ⁵	106.92	144.71	104.40	149.53	98.36	156.69	92.01	167.11
127	111.18	140.85	107.63	146.16	104.58	150.19	99.09	158.25	92.59	167.12
128	111.98	142.14	108.91 ⁵	147.94	105.90 ⁵	150.36	100.61	159.53	94.36	168.52
129	113.21	143.49	110.26	147.94	107.32	151.63	102.16 ⁵	159.67	95.01 ⁵	170.14
130	114.47 ⁵	145.01	111.75	148.35	109.11	152.96	102.16 ⁵	161.01	95.01 ⁵	171.24
131	115.79	146.45	113.35	149.60	109.61	154.39	102.42	162.46	96.40	171.81
132	117.19	146.45	113.35	150.88	109.61	156.32	103.84	164.31	97.95	173.28 ⁵
133	118.95	146.66	113.35	152.21	110.11	156.32	105.66	164.31	97.95	175.08
134	119.50	147.86	114.11	153.67	111.44	156.87	106.12	165.33	98.59	175.11
135	119.50	149.08	115.40	155.43	112.87	158.15	106.12	166.71	100.36	176.51
136	119.50	150.33	116.74	155.43	114.84	159.48	107.10	168.29	101.02	178.11
137	120.44	151.61	118.21	155.81 ⁵	114.84	160.92 ⁵	108.61 ⁵	169.49	101.02	179.27
138	121.67	152.96	119.88	157.05	114.84	162.79	110.16	169.64	102.34	179.76 ⁵
139	122.92 ⁵	154.45	119.88	158.33	115.60 ⁵	162.79	110.16	170.98	104.01	181.21
140	124.23	156.01	119.88	159.65	116.93	163.35	110.37	172.41	104.01	183.15
141	125.61	156.01	120.55	161.07 ⁵	118.35	164.63	111.78	174.36	104.50	183.15
142	127.22	156.09	121.82	163.02	120.36	165.96	113.45	174.36	106.13	184.41
143	128.23	157.28	123.15	163.02	120.36	167.39	114.33	175.25	107.23 ⁵	185.94
144	128.23	158.49	124.57	163.24	120.36	169.33	114.33	176.61	107.23 ⁵	187.43
145	128.23	159.73	126.52	164.47	121.06	169.33	114.99	178.11	108.17	187.65
146	128.84	160.99 ⁵	126.52	165.73	122.37	169.80	116.44	179.67	110.17	189.06
147	130.06	162.31	126.52	167.03	123.77	171.07	118.33	179.67	110.17	190.71
148	131.30	163.72 ⁵	126.94 ⁵	168.41	125.46	172.38 ⁵	118.33	180.84	110.34	191.67
149	132.57 ⁵	165.68	128.20	170.00	126.26	173.79	118.33	182.22	111.85	192.25
150	133.91	165.77	129.50	171.09	126.26	175.48 ⁵	119.59	183.81	113.51	193.70

If c is the observed frequency or count and m_L, m_U are the lower and upper $100(1-\epsilon)\%$ confidence limits for its expectation, m , then $\Pr(m_L \leq m \leq m_U) \geq 1-\epsilon$.

Confidence coefficient, $100(1-\epsilon)\%$

c	80		90		95		99		99.9	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
151	135.37	165.77	130.87	171.09	126.48	176.23	121.09	184.97 ⁵	113.51	195.67
152	137.08	166.64 ⁵	132.44	171.84	127.78	176.23	122.69	185.08	113.95	195.67
153	137.08	167.84	133.66	173.08	129.14	177.48	122.69	186.40	115.52	196.84
154	137.08	169.06	133.66	174.36	130.68	178.77	122.78	187.81	116.84	198.33
155	137.18	170.30 ⁵	133.66	175.69	132.03	180.14	124.16	189.50	116.84	200.04
156	138.38	171.58	134.54	177.13	132.03	181.67	125.70	190.28	117.54	200.05
157	139.60 ⁵	172.92	135.81	178.96	132.03	183.05	127.07	190.61 ⁵	119.21	201.42
158	140.85	174.39	137.13	178.96	133.14 ⁵	183.05	127.07	191.94	120.17	202.94
159	142.14	176.06	138.55	179.18	134.48	183.86	127.31	193.36	120.17	204.43
160	143.49	176.06	140.54	180.40 ⁵	135.92	185.13	128.70	195.19	121.12	204.59
161	145.01	176.06	140.54	181.65 ⁵	137.79	186.46	130.27 ⁵	195.59	122.92	205.98
162	146.45	177.14	140.54	182.94	137.79	187.89	131.50	196.13	123.46	207.55
163	146.45	178.34	140.85	184.30	137.79	189.83	131.50	197.46	123.46	208.83
164	146.45	179.56	142.09	185.80	138.49	189.83	131.82	198.88	124.68	209.13
165	146.66	180.80	143.37	187.30	139.79	190.21	133.21	200.84	126.67	210.52
166	147.86	182.08	144.71	187.30	141.16	191.46	134.79	200.94	126.67	212.12
167	149.08	183.42	146.16	187.70	142.70	192.76	135.99	201.62	126.75	213.26
168	150.33	184.89	147.94	188.93	144.01	194.11 ⁵	135.99	202.94	128.21	213.64
169	151.61	186.56 ⁵	147.94	190.18	144.01	195.63	136.30	204.36	130.14	215.04
170	152.96	186.56 ⁵	147.94	191.47	144.01	197.09	137.68	206.19	130.14	216.66 ⁵
171	154.45	186.56 ⁵	148.35	192.83	145.08	197.09	139.24	206.60	130.26	217.72
172	156.01	187.59	149.60	194.36	146.39	197.78	140.54	207.08	131.72	218.14
173	156.01	188.78	150.88	195.73	147.80	199.04	140.54	208.40	133.64	219.54
174	156.01	189.99	152.21	195.73	149.53	200.35	140.76	209.81	133.64	221.16
175	156.09	191.23	153.67	196.18	150.19	201.73	142.12	211.50	133.75	222.22
176	157.28	192.49	155.43	197.40 ⁵	150.19	203.35 ⁵	143.64	212.29	135.20	222.63
177	158.49	193.81	155.43	198.65 ⁵	150.36	204.36	145.13	212.53	137.15	224.02
178	159.73	195.22	155.43	199.94	151.63	204.36	145.13	213.84	137.15	225.62
179	160.99 ⁵	197.12	155.81 ⁵	201.30	152.96	205.31 ⁵	145.19	215.22	137.22	226.76
180	162.31	197.33	157.05	202.80	154.39	206.58	146.54	216.80	138.66	227.09 ⁵
181	163.72 ⁵	197.33	158.33	204.30	156.32	207.90	148.01	217.98	140.64	228.48
182	165.68	197.97	159.65	204.30	156.32	209.30	149.76	217.98	140.74	230.04
183	165.77	199.16	161.07 ⁵	204.62	156.32	211.03	149.76	219.25	140.74	231.34
184	165.77	200.36	163.02	205.84	156.87	211.69	149.76	220.61	142.10	231.55
185	165.77	201.58	163.02	207.08	158.15	211.69	150.93	222.10 ⁵	143.83	232.92
186	166.64 ⁵	202.82 ⁵	163.02	208.36	159.48	212.82	152.35 ⁵	223.67 ⁵	144.55	234.43 ⁵
187	167.84	204.11	163.24	209.69	160.92 ⁵	214.09	154.18	223.67 ⁵	144.55	235.94
188	169.06	205.45	164.47	211.13	162.79	215.40	154.60	224.65	145.51 ⁵	235.99
189	170.30 ⁵	206.94	165.73	212.96 ⁵	162.79	216.81	154.60	225.98	147.13 ⁵	237.34
190	171.58	208.52	167.03	212.96 ⁵	162.79	218.56	155.31	227.41	148.24	238.82
191	172.92	208.52	168.41	213.03	163.35	219.16	156.69	229.37	148.24	240.56
192	174.39	208.52	170.00	214.24	164.63	219.16	158.25	229.37	148.92	240.56
193	176.06	209.49	171.09	215.47	165.96	220.29	159.53	230.03	150.46	241.75
194	176.06	210.67 ⁵	171.09	216.73	167.39	221.56	159.53	231.33	151.91	243.19
195	176.06	211.88	171.09	218.03	169.33	222.86 ⁵	159.67	232.71	151.91	245.18
196	176.06	213.10	171.84	219.40 ⁵	169.33	224.26	161.01	234.28	152.31	245.18
197	177.14	214.35	173.08	221.00	169.33	225.90 ⁵	162.46	235.50	153.79	246.15
198	178.34	215.63	174.36	222.10	169.80	226.81	164.31	235.50	155.57 ⁵	247.55
199	179.56	216.98 ⁵	175.69	222.10	171.07	226.81	164.31	236.68	155.57 ⁵	249.18
200	180.80	218.49	177.13	222.60	172.38 ⁵	227.73	164.31	238.01	155.69	250.20

THE DISTRIBUTION OF THE NUMBER OF SUCCESSES IN A SEQUENCE OF DEPENDENT TRIALS

BY K. R. GABRIEL

Hebrew University, Jerusalem and University of North Carolina

1. INTRODUCTION

Consider a sequence of n trials and an initial trial whose outcomes are the random variables $X_1, X_2, \dots, X_n$ and X_0 respectively, where

$$X_k \begin{cases} = 1 & \text{if the } k\text{th trial is successful} \\ = 0 & \text{otherwise,} \end{cases}$$

for $k = 0, 1, 2, \dots, n$. Let the probability of a success at the initial trial be

$$R = \Pr \{X_0 = 1\},$$

and assume the conditional probabilities, which are denoted by

$$p_1 = \Pr \{X_k = 1 \mid X_{k-1} = 1\},$$

$$p_0 = \Pr \{X_k = 1 \mid X_{k-1} = 0\},$$

for $k = 1, 2, \dots, n$, to be independent of k and of the outcomes of the trials preceding the $(k-1)$ st.

This sequence is an irreducible Markov Chain with two ergodic states. Its stationary probability distribution has as probability of a success

$$P = \frac{p_0}{1 - (p_1 - p_0)}.$$

These results follow from the general theory of Markov Chains (Feller, 1957, chapters xv and xvi.2, example (a)).

The recurrence time distribution of successes has probabilities

$$p_1, (1-p_1)p_0, (1-p_1)(1-p_0)p_0, (1-p_1)(1-p_0)^2p_0, \dots,$$

for one, two, three, four, ... trials up to and including the next success. Hence the mean and variance of recurrence times are found to be

$$\mu = \frac{1 - (p_1 - p_0)}{p_0}, \quad \sigma^2 = \frac{1 - p_1}{p_0^2} (1 + p_1 - p_0).$$

Now consider the number of successes in the sequence, i.e.

$$S = \sum_{k=1}^n X_k.$$

It is known that S is asymptotically normally distributed with mean n/μ and variance $n\sigma^2/\mu^3$ (Feller, 1957, chapter XIII.6; Smith, 1958). Substituting for μ and σ^2 , one obtains

$$E(S) \sim nP, \quad \text{Var}(S) \sim nP(1-P) \frac{1+d}{1-d},$$

where $d = p_1 - p_0$. This is an asymptotic result which indicates neither the exact distribution for small n nor the rapidity of approach to normality. It seems to have been given first by Markov (1924, § 60).

2. THE EXACT DISTRIBUTION

If S successes occur in n trials there will be a number of changes from success on one trial (including the initial trial) to failure on the next, and vice versa. Denote the number of changes by C and let a and b stand for the least integer not smaller than $\frac{1}{2}C - 1$ or $\frac{1}{2}C$ respectively.

Consider first the case of a success at the initial trial, then S successes with C changes will involve b changes from success to failure, a changes from failure to success, as well as $S - a$ successes following successes and $n - S - b$ failures following failures. The probability of any one arrangement of S successes and $n - S$ failures with C changes is therefore

$$(1 - p_1)^b p_0^a p_1^{S-a} (1 - p_0)^{n-S-b},$$

i.e.

$$p_1^S (1 - p_0)^{n-S} \left(\frac{1 - p_1}{1 - p_0} \right)^b \left(\frac{p_0}{p_1} \right)^a.$$

Next, any arrangement of n trials with S successes and C changes will involve a changes to successes which may occur before any a of the S successes, i.e. in any of $\binom{S}{a}$ different possible positions. Also b changes occur before failures, of which the first must occur before the first failure and the rest may be arranged in any of $\binom{n-S-1}{b-1}$ different ways. Given the arrangements of both kinds of changes, the position of all changes in the sequence is uniquely determined, and therefore the total number of possible arrangements of C changes among n trials with S successes is $\binom{S}{a} \binom{n-S-1}{b-1}$. Hence the probability of S successes with C changes in n trials following a successful initial trial is

$$\Pr \{S, C \mid n, X_0 = 1\} = \binom{S}{a} \binom{n-S-1}{b-1} p_1^S (1 - p_0)^{n-S} \left(\frac{1 - p_1}{1 - p_0} \right)^b \left(\frac{p_0}{p_1} \right)^a.$$

The number of changes C may be any number from 1 to $n + \frac{1}{2} - |2S - \frac{1}{2} + n| = C_1$, say, except if $S = n$ when $C_1 = 0$. Thus the probability of S successes in n trials following a successful initial trial is obtained by summing the above probabilities over all possible values of C ,

$$\Pr \{S \mid n, X_0 = 1\} = p_1^S (1 - p_0)^{n-S} \sum_{C=1}^{C_1} \binom{S}{a} \binom{n-S-1}{b-1} \left(\frac{1 - p_1}{1 - p_0} \right)^b \left(\frac{p_0}{p_1} \right)^a.$$

In a similar manner for the case of a failure on the initial trial

$$\Pr\{S | n, X_0 = 0\} = p_1^S (1-p_0)^{n-S} \sum_{C=1}^{C_0} \binom{S-1}{b-1} \binom{n-S}{a} \left(\frac{1-p_1}{1-p_0}\right)^a \left(\frac{p_0}{p_1}\right)^b,$$

where $C_0 = n + \frac{1}{2} - |2S - \frac{1}{2} - n|$, except for $S = 0$ when $C_0 = 0$ and the summation above is understood to include only the term $C = 0$.

As one would expect, when $d = 0$, i.e. in the case of independent trials, either probability becomes that of the binomial distribution.

Finally, given the probability R of a success at the initial trial, the unconditional probability of S successes in the sequence of n trials becomes

$$\Pr\{S | n\} = R \cdot \Pr\{S | n, X_0 = 1\} + (1-R) \cdot \Pr\{S | n, X_0 = 0\}.$$

This form of the distribution is rather unwieldy but does not seem to lend itself to simplification. It may be easier to obtain an impression of the closeness to normality by means of the moments.

3. MOMENTS OF THE DISTRIBUTION

For a sequence of n dependent trials the r th factorial moment of the number of successes S is

$$r! \phi_r = r! \sum_{i_1 < i_2 < \dots < i_r} \Pr\{X_{i_1} = 1, X_{i_2} = 1, \dots, X_{i_r} = 1\},$$

where $i_1, i_2, \dots, i_r$ is any subset of r of the integers $1, 2, \dots, n$ (Gumbel, 1938, table 1 and formula 17). Putting $k_j = i_j - i_{j-1}$ ($j = 2, \dots, r$) and $k_1 = i_1$ this becomes

$$r! \phi_r = r! \sum^{(r)} \Pr\{X_{k_1} = 1, X_{k_1+k_2} = 1, \dots, X_{k_1+k_2+\dots+k_r} = 1\},$$

where $\sum^{(r)}$ denotes summation over all sets of r positive integer suffices $(k_1, \dots, k_r)$ whose sum is $\leq n$. For a Markov Chain this becomes

$$r! \sum^{(r)} \Pr\{X_{k_1} = 1\} \Pr\{X_{k_1+k_2} = 1 | X_{k_1} = 1\} \dots \Pr\{X_{k_1+k_2+\dots+k_r} = 1 | X_{k_1+k_2+\dots+k_{r-1}} = 1\}.$$

Now $\Pr\{X_k = 1 | X_{k-i} = \gamma\} = P(1-d^{i-1}) + d^{i-1} \Pr\{X_{k-i+1} = 1 | X_{k-i} = \gamma\}$

for $k = 1, 2, \dots, n$ and $i = 1, 2, \dots, k$ (Feller, 1957, p. 385). In particular,

$$\Pr\{X_{k_1+k_2+\dots+k_i} = 1 | X_{k_1+k_2+\dots+k_{i-1}} = 1\} = P + (1-P) d^{k_i}$$

and

$$\Pr\{X_{k_1} = 1\} = P + (R-P) d^{k_1}.$$

Introducing these probabilities into the expression for the factorial moments, and writing $f = 1/(1-d)$, one obtains

$$\phi_1 = nP + (R-P) df(1-d^n),$$

$$\phi_2 = P^2 \frac{n^2 - n}{2} + P(1+R-2P) df^2[n(1-d) - 1(1-d^n)] + (1-P)(R-P) df^2[d(1-d^n) - nd^n(1-d)],$$

$$\begin{aligned} \phi_3 = & P^3 \frac{n^3 - 3n^2 + 2n}{6} + P^2(2+R-3P) \frac{1}{2} df^3[n^2(1-d)^2 - n(1-d)(3-d) + 2(1-d^n)] \\ & + P(1-P)(1+2R-3P) df^3[nd(1-d) - 2d(1-d^n) + nd^n(1-d)] \\ & + (R-P)(1-P)^2 df^3[d^2(1-d^n) - \frac{1}{2} n^2 d^n(1-d)^2 + \frac{1}{2} nd^n(1-d)(1-3d)], \end{aligned}$$

$$\begin{aligned} \phi_4 = & P^4 \frac{n^4 - 6n^3 + 11n^2 - 6n}{24} + P^3(3 + R - 4P) \frac{1}{6} df^4 [n^3(1-d)^3 - 3n^2(1-d)^2(2-d) \\ & + n(1-d)(11-7d+2d^2) - 6(1-d^n)] + 3P^2(1-P)(1+R-2P) \\ & \times \frac{1}{2} df^4 [n^2d(1-d)^2 - nd(1-d)(5-d) - 2nd^n(1-d) + 6d(1-d^n)] \\ & + P(1-P)^2(1+3R-4P) df^4 [nd^2(1-d) - 3d^2(1-d^n) + \frac{1}{2}n^2d^n(1-d)^2 \\ & - \frac{1}{2}nd^n(1-d)(1-5d)] + (R-P)(1-P)^3 df^4 [d^3(1-d^n) - \frac{1}{6}nd^n(1-d)(2-7d+11d^2) \\ & + \frac{1}{2}n^2d^n(1-d)^2(1-2d) - \frac{1}{6}n^3d^n(1-d)^3]. \end{aligned}$$

These expressions are unwieldy and their substitution into the formulae for the moments about the mean (Kendall, 1948, chapter 3) does not yield simpler forms. In numerical work it would appear advisable to compute the ϕ values first and then obtain the other moments from them.

Again, for $d = 0$ these moments become those of the binomial distribution.

In the special case $R = P$, whatever d , i.e. in a stationary Markov Chain, the first three moments are:

$$\mu_1 = nP,$$

$$\mu_2 = P(1-P)[n(1+d)f - 2df^2(1-d^n)],$$

$$\mu_3 = P(1-P)(1-2P)[n\{1 + 6df^2(1+d^n)\} - 6d(1+d)f^3(1-d^n)].$$

For large, n , approximations may be obtained by ignoring terms in d^n and terms not involving n .* Then

$$\mu_1 \approx nP,$$

$$\mu_2 \approx nP(1-P) \frac{1+d}{1-d},$$

$$\gamma_1 \approx \frac{1}{\sqrt{n}} \frac{1-2P}{\sqrt{[P(1-P)]}} \frac{1+4d+d^2}{(1+d)\sqrt{(1-d^2)}},$$

$$\gamma_2 \approx \frac{1}{n} \frac{1-6P(1-P)}{P(1-P)} \frac{1+10d+d^2}{1-d^2},$$

where terms involving $(R-P)$ and n^{-1} or powers of n^{-1} are ignored, and γ_1, γ_2 are the usual standardized 3rd and 4th cumulants.

At this level of approximation the cumulants are independent of R , i.e. for a sufficiently long sequence the outcome of the initial trial is immaterial. μ_2, γ_1 and γ_2 can be factorized into the corresponding term for the binomial distribution and a term depending on d . For large n , γ_1 and γ_2 will vanish whilst μ_1 and μ_2 correspond to the mean and variance in the asymptotic normal distribution (§1). For any n, P and d , the approximations to γ_1 and γ_2 may give some indication of the deviation from normality.

* [By introducing the moments of the recurrence time distribution, these large sample approximations to the moments of S can presumably be related to the results of W. L. Smith, whose paper (*Biometrika*, 1959, 46, 1-29) 'On the cumulants of renewal processes' was published after the present paper had been received for publication. It should be noted, however, that while Smith's results were established for the continuous case, the present author deals with a discrete chain.—ED.]

4. NUMERICAL CALCULATIONS AND AN APPLICATION

An idea of the form of the distribution of S is given by the cumulants as computed for Tables 1 and 2. The exact computations are cumbersome and have therefore been restricted to a few selected values of the parameters. It will be noted that for values of $P = 0.2$ and 0.4 , μ_2 and γ_2 will be as for $P = 0.8$ and 0.6 , respectively, γ_1 will change sign, and the mean will be n minus the mean for $P = 0.8$ and 0.6 .

Table 1. Cumulants of the distribution of S in sequences of seven trials

$P \dots$		0.6				0.8			
$d \dots$		0	0.2	0.4	0.8	0	0.2	0.4	0.8
μ_1		4.2	4.2	4.2	4.2	5.6	5.6	5.6	5.6
		1.68	2.37	3.39	7.53	1.12	1.58	2.26	5.02
μ_2	Exact	1.68	2.37	3.39	7.53	1.12	1.58	2.26	5.02
	Approx.	1.68	2.52	3.92	15.12	1.12	1.68	2.61	10.08
γ_1	Exact	-0.154	-0.228	-0.294	-0.394	-0.567	-0.837	-1.081	-1.447
	Approx.	-0.154	-0.241	-0.332	-0.691	-0.567	-0.887	-1.219	-2.540
γ_2	Exact	-0.261	-0.442	-0.690	-1.425	0.036	0.380	0.686	0.709
	Approx.	-0.261	-0.828	-1.606	-7.001	0.036	0.113	0.219	0.956

Table 2. Cumulants of the distribution of S in sequences of thirty trials

$P \dots$		0.6				0.8			
$d \dots$		0	0.2	0.4	0.8	0	0.2	0.4	0.8
μ_1		18	18	18	18	24	24	24	24
		7.20	10.65	16.27	55.45	4.80	7.10	10.84	36.81
μ_2	Exact	7.20	10.65	16.27	55.45	4.80	7.10	10.84	36.81
	Approx.	7.20	10.80	16.80	64.80	4.80	7.20	11.19	43.20
γ_1	Exact	-0.075	-0.115	-0.157	-0.297	-0.274	-0.423	-0.576	-1.099
	Approx.	-0.075	-0.117	-0.160	-0.334	-0.274	-0.429	-0.589	-1.227
γ_2	Exact	-0.061	-0.109	-0.181	-0.684	0.008	0.105	0.227	0.771
	Approx.	-0.061	-0.193	-0.375	-1.633	0.008	0.026	0.051	0.223

For $d = 0$ the distribution is binomial, as was noted. With increasing d , the variance, skewness and kurtosis increase considerably. Changes in n and P affect the moments much as they do those of the binomial. This regularity holds only if d^n is negligible, e.g. for $d = 8$ and $n = 7$ the pattern is disrupted.

The closeness of the approximations seems to depend largely on d being small and n large. The approximations are a good deal better for μ_2 and γ_1 than for γ_2 .

The distribution of S is illustrated by an application to rainfall data of Tel Aviv, where it has been suggested that a Markov Chain probability model holds* (rainy and dry days

* The fit of a Markov Chain probability scheme to daily rainfall data for Tel Aviv will be discussed in a forthcoming publication by K. R. Gabriel and J. Neumann.

being considered as 'successes' and 'failures', respectively). The estimated probabilities for January are $p_1 = 0.674$ and $p_0 = 0.293$.

The exact cumulants are therefore

$$\mu_1 = 14.663, \quad \mu_2 = 16.744, \quad \gamma_1 = +0.040, \quad \gamma_2 = -0.195,$$

while the large sample approximate value of the variance is

$$\mu_2 = nP(1-P)(1+d)/(1-d) = 17.240.$$

Table 3 shows that for these parameter values the normal approximation using the large sample variance is very close to the discrete exact distribution. Observed data are available for only 27 years and are therefore not presented in the table.

Table 3. *Exact distribution and asymptotic normal distribution, the comparison being made for the values $n = 31$, $P = 0.473$, $d = 0.381$ (suggested by data of January rainfall at Tel Aviv)*

No. of successes	Probability		No. of successes	Probability		No. of successes	Probability	
	Exact	Approx.		Exact	Approx.		Exact	Approx.
0	0.0000	0.0003				21	0.0306	0.0301
1	0.0001	0.0004	11	0.0671	0.0651	22	0.0206	0.0202
2	0.0004	0.0009	12	0.0794	0.0782	23	0.0129	0.0129
3	0.0010	0.0019	13	0.0890	0.0886	24	0.0075	0.0078
4	0.0027	0.0036	14	0.0942	0.0947	25	0.0040	0.0044
5	0.0057	0.0065	15	0.0946	0.0956			
						26	0.0020	0.0023
6			16	0.0899	0.0907	27	0.0009	0.0012
7	0.0105	0.0110	17	0.0810	0.0819	28	0.0003	0.0006
8	0.0177	0.0176	18	0.0692	0.0696	29	0.0001	0.0002
9	0.0275	0.0266	19	0.0558	0.0557	30	0.0000	0.0001
10	0.0396	0.0380	20	0.0426	0.0422	31	0.0000	0.0001
	0.0531	0.0510						

Note on computation

Computation of the exact probability distribution is rather tedious. It involves evaluation of the $n(n+1)+2$ terms inside the summation signs of the expressions for

$$\Pr\{S | n, X_0 = 1\} \quad \text{and} \quad \Pr\{S | n, X_0 = 0\}.$$

This is best done with the help of tables of logarithms of binomial coefficients (Hald, 1952, table XIV). I am happy to acknowledge Mrs R. Falk's precise work in computing the distribution of Table 3.

I wish to thank Profs. W. J. Hall and W. Hoeffding for their helpful suggestions.

REFERENCES

- FELLER, W. (1957). *An Introduction to Probability Theory and its Applications*. **1** (second edition). New York: John Wiley and Sons.
- GUMBEL, E. V. (1938). Gli eventi compatibili. *Giorn. Ist. It. Att.* **9**, 58-93.
- HALD, A. (1952). *Statistical Tables and Formulas*. New York: John Wiley and Sons.
- KENDALL, M. G. (1948). *The Advanced Theory of Statistics*. **1**. London: Charles Griffin and Co. Ltd.
- MARKOV, A. A. (1924). *Probability Theory* (4th ed.). Moscow (in Russian).
- SMITH, W. L. (1958). *Lectures on Stochastic Processes at the University of North Carolina*. (Unpublished.)

Editorial note added in proof

The reader may care to note that a paper by P. V. Krishna Iyer and N. S. Shakuntala on 'Cumulants of some distributions arising from a two-state Markoff chain', which has appeared since the present paper was sent to press (1959, *Proc. Camb. Phil. Soc.* **55**, 273-6) covers some ground in parallel with that covered by Mr Gabriel.

E. S. P.

TABLE OF THE UPPER 10% POINTS OF THE STUDENTIZED RANGE

BY J. PACHARES

Hughes Aircraft Company, Culver City, California

1. INTRODUCTION

Let $q = w/s$ be the familiar 'studentized' range when sampling from a normal population, where w is a sample range of n values and s is a sample estimate of standard deviation based on ν degrees of freedom and distributed independently of w . If $F(q) = \Pr(w/s \leq q)$, then the values of q for which $F(q) = 1 - \alpha$ have been computed to three significant figures for $\alpha = 0.10$, $n = 2$ (1) 20 and $\nu = 1$ (1) 20, 24, 30, 40, 60, 120, ∞ . The values of q when $\alpha = 0.05$ were also recomputed for checking purposes with the result that several minor disagreements with Pearson & Hartley's table 29 (1954) were noted.* The need for 10 % points as well as more extreme points was originally suggested by Prof. Henry Scheffé (1953, p. 91) in his paper discussing methods for judging all contrasts in the analysis of variance.

2. FORMULAE AND NOTATION

Let
$$z(x) = (2\pi)^{-\frac{1}{2}} e^{-x^2/2}, \quad E(t) = \int_0^t z(x) dx, \quad c_\nu = 2\{\frac{1}{4}\sqrt{(\pi\nu)}\}^\nu / \Gamma(\frac{1}{2}\nu)$$

and let $P_n(t)$ be the cumulative distribution function of the range in samples of n from a $N(0, 1)$ population. Using the form given by May (1952), we have that

$$1 - F(q) = c_\nu \int_0^\infty \{4(t/q) z(t/q)\}^\nu (1 - P_n(t)) t^{-1} dt, \quad (1)$$

where

$$P_n(t) = n \int_{-\infty}^\infty z(u) (E(u+t) - E(u))^{n-1} du. \quad (2)$$

An alternate form for $P_n(t)$ derived by Hartley (1942) is given by

$$P_n(t) = (2E(\frac{1}{2}t))^n + 2n \int_{\frac{1}{2}t}^\infty z(u) (E(u) - E(u-t))^{n-1} du. \quad (3)$$

If $n = 2$ it is easy to show that

$$F(q) = \frac{2}{B(\frac{1}{2}, \frac{1}{2}\nu)} \int_0^{\arctan(q/\sqrt{(2\nu)})} (\cos u)^{\nu-1} du. \quad (4)$$

3. METHOD OF COMPUTATION

A subroutine was first programmed using the IBM 704 computer to evaluate $E(t)$ to at least seven significant figures by storing a table of $E(t)$ values in memory for $t = 0$ (0.2) 3 (0.4) 7 (1) 9, then using the first six terms of the Taylor series expansion of $E(t)$ about the nearest value of t in the stored table. Two routines, herein referred to as routine 1 and routine 2, were then programmed for the IBM 704, both making use of the $E(t)$ sub-

* See Editorial Note on p. 462 below.

routine already mentioned. Routine 1 was used to solve for the percentage points and routine 2 was used for checking purposes. Routine 1 computed $P_n(t)$ from (3) for $t = 0(0.125)8$ and $n = 2(1)20$, these values being stored on magnetic tape for use in (1). The values of q which made $F(q) = 1 - \alpha$ were found to three significant figures by iteration using linear interpolation in q . The integral in (3) was computed by the Euler-Maclaurin integration formula with $u = \frac{1}{2}t(0.125)8$ using terms up to and including the first derivative of the integrand at the lower limit; the derivative at the upper limit was taken to be zero. The integral in (1) was computed using the trapezoid rule with $t = 0(0.125)8$. The values of q when $n = 2$ and $\nu = 1, 2$ were found explicitly using (4). For $\nu = \infty$, the values of q were computed *ab initio* using iteration and linear interpolation to solve the equation $P_n(q) = 1 - \alpha$, where $P_n(q)$ was computed from (2) using the rectangular rule with $u = -4.5(0.125)4.5$. Starting values were found for the first column of the table (i.e. $n = 2$) by using a crude interpolation between the values found at $\nu = 1, 2$ and $\nu = \infty$. The solution of a given case served as the initial guess for the following case.

4. METHOD OF CHECKING

The results of routine 1 were checked by substituting them into routine 2 which computed $P_n(t)$ from (2) using the rectangular rule with $u = -6(0.375)6$. Routine 2 evaluated the integral in (1) using the Euler-Maclaurin integration formula with terms up to and including the first derivative of the integrand at the lower limit of integration, the value at the upper limit being taken to be zero, the integration being stopped when the value of the integrand became less than 10^{-8} . The checking process resulted in altering only the first few values on the $\nu = 2$ row.

The writer wishes to express his sincere appreciation to Lois Matsunaga for her patient and skilful programming of the routines needed to produce the final results. Thanks are also due to Prof. Henry Scheffé for indicating the need for the upper 10% points of the 'Studentized' range.

REFERENCES

- HARTLEY, H. O. (1942). *Biometrika*, **32**, 334-48.
 MAY, J. M. (1952). *Biometrika*, **39**, 192-3.
 PEARSON, E. S. & HARTLEY, H. O. (1954). *Biometrika Tables for Statisticians*, **1**. Cambridge University Press.
 SCHEFFÉ, H. (1953). *Biometrika*, **40**, 87-104.

CORRECTED TABLES OF UPPER 5% AND 1% POINTS OF THE STUDENTIZED RANGE

Editorial note

A short explanation of the basis of the accompanying tables of 10%, 5% and 1% points of the studentized range seems called for. In publishing the table of 5% and 1% points of this ratio, based on Miss Joyce May's calculations, Prof. Hartley and I (*Biometrika Tables for Statisticians*, **1**, 1954, table 29, and introduction, p. 52) recognized that a certain number of last figure errors were likely to be present, which we hoped would not exceed one unit. In his calculation of the 10% points described in the preceding paper Dr James Pachares recomputed the 5% values for checking purposes and was thus able to identify about 40 unit errors in our last figure.

After receiving Dr Pachares's paper, we learned that some very extensive computations on the distribution of range and studentized range were being carried out under the direction of Dr Leon Harter in the Aeronautical Research Laboratory of the Wright Air Development Center, Ohio. This

work is being published as a *W.A.D.C. Technical Report* (no. 58-484, vols. I and II). Since this programme included the calculation of 10 %, 5 % and 1 % points of the studentized range to high accuracy, the editors of *Biometrika* felt that it would be useful to take this opportunity of printing, with Dr Pachares's paper, corrected tables of both the upper 5 % and 1 % points, given throughout to two decimal places.

We are most grateful to Dr Harter for allowing *Biometrika* to make use of his 1 % results in advance of publication. This has made it possible to identify some 30 last figure unit errors in Miss May's original 1 % table. Further, in the tables at all three levels it has been possible to include a second decimal for low values of ν , where both Miss May and Dr Pachares had rounded off to one decimal place only*. Finally, Dr Harter and his collaborators have gone to the trouble of recomputing certain values with a smaller tolerance where a 5 in the final figure of their fuller tables left uncertainty in the rounding off to two decimal places.

An interesting point which appeared in a comparison of the Pachares and Harter rules for determining the final figure in rounding off for the 100 P % point, was that under the former rule that value of x is taken which minimizes $|F(x) - P|$, i.e. gives a probability integral nearest to the required value of P . On the other hand the Harter rule takes for x that value which is nearest to the value of X for which $F(X) = P$ exactly. These rules almost invariably lead to the same answer, since $F(x)$ will increase nearly linearly for small increments of 0.01 in x . However, in one case at least the rules do lead to different conclusions: for the 10% point, $n = 12$, $\nu = 16$ the Pachares rule gives 4.81 and the Harter rule 4.80; the former value has been inserted in Table 1.

E. S. P.

* Dr Harter has asked me to make clear, however, that where the tabular value lies between 10 and 100 and two decimals are recorded he cannot guarantee that the second decimal place is not in error by one unit.

Table 1. *Upper 10% points of the studentized range*

$\nu \backslash n$	2	3	4	5	6	7	8	9	10
1	8.93	13.44	16.36	18.49	20.15	21.51	22.64	23.62	24.48
2	4.13	5.73	6.77	7.54	8.14	8.63	9.05	9.41	9.72
3	3.33	4.47	5.20	5.74	6.16	6.51	6.81	7.06	7.29
4	3.01	3.98	4.59	5.03	5.39	5.68	5.93	6.14	6.33
5	2.85	3.72	4.26	4.66	4.98	5.24	5.46	5.65	5.82
6	2.75	3.56	4.07	4.44	4.73	4.97	5.17	5.34	5.50
7	2.68	3.45	3.93	4.28	4.55	4.78	4.97	5.14	5.28
8	2.63	3.37	3.83	4.17	4.43	4.65	4.83	4.99	5.13
9	2.59	3.32	3.76	4.08	4.34	4.54	4.72	4.87	5.01
10	2.56	3.27	3.70	4.02	4.26	4.47	4.64	4.78	4.91
11	2.54	3.23	3.66	3.96	4.20	4.40	4.57	4.71	4.84
12	2.52	3.20	3.62	3.92	4.16	4.35	4.51	4.65	4.78
13	2.50	3.18	3.59	3.88	4.12	4.30	4.46	4.60	4.72
14	2.49	3.16	3.56	3.85	4.08	4.27	4.42	4.56	4.68
15	2.48	3.14	3.54	3.83	4.05	4.23	4.39	4.52	4.64
16	2.47	3.12	3.52	3.80	4.03	4.21	4.36	4.49	4.61
17	2.46	3.11	3.50	3.78	4.00	4.18	4.33	4.46	4.58
18	2.45	3.10	3.49	3.77	3.98	4.16	4.31	4.44	4.55
19	2.45	3.09	3.47	3.75	3.97	4.14	4.29	4.42	4.53
20	2.44	3.08	3.46	3.74	3.95	4.12	4.27	4.40	4.51
24	2.42	3.05	3.42	3.69	3.90	4.07	4.21	4.34	4.44
30	2.40	3.02	3.39	3.65	3.85	4.02	4.16	4.28	4.38
40	2.38	2.99	3.35	3.60	3.80	3.96	4.10	4.21	4.32
60	2.36	2.96	3.31	3.56	3.75	3.91	4.04	4.16	4.25
120	2.34	2.93	3.28	3.52	3.71	3.86	3.99	4.10	4.19
∞	2.33	2.90	3.24	3.48	3.66	3.81	3.93	4.04	4.13

$\nu \backslash n$	11	12	13	14	15	16	17	18	19	20
1	25.24	25.92	26.54	27.10	27.62	28.10	28.54	28.96	29.35	29.71
2	10.01	10.26	10.49	10.70	10.89	11.07	11.24	11.39	11.54	11.68
3	7.49	7.67	7.83	7.98	8.12	8.25	8.37	8.48	8.58	8.68
4	6.49	6.65	6.78	6.91	7.02	7.13	7.23	7.33	7.41	7.50
5	5.97	6.10	6.22	6.34	6.44	6.54	6.63	6.71	6.79	6.86
6	5.64	5.76	5.87	5.98	6.07	6.16	6.25	6.32	6.40	6.47
7	5.41	5.53	5.64	5.74	5.83	5.91	5.99	6.06	6.13	6.19
8	5.25	5.36	5.46	5.56	5.64	5.72	5.80	5.87	5.93	6.00
9	5.13	5.23	5.33	5.42	5.51	5.58	5.66	5.72	5.79	5.85
10	5.03	5.13	5.23	5.32	5.40	5.47	5.54	5.61	5.67	5.73
11	4.95	5.05	5.15	5.23	5.31	5.38	5.45	5.51	5.57	5.63
12	4.89	4.99	5.08	5.16	5.24	5.31	5.37	5.44	5.49	5.55
13	4.83	4.93	5.02	5.10	5.18	5.25	5.31	5.37	5.43	5.48
14	4.79	4.88	4.97	5.05	5.12	5.19	5.26	5.32	5.37	5.43
15	4.75	4.84	4.93	5.01	5.08	5.15	5.21	5.27	5.32	5.38
16	4.71	4.81	4.89	4.97	5.04	5.11	5.17	5.23	5.28	5.33
17	4.68	4.77	4.86	4.93	5.01	5.07	5.13	5.19	5.24	5.30
18	4.65	4.75	4.83	4.90	4.98	5.04	5.10	5.16	5.21	5.26
19	4.63	4.72	4.80	4.88	4.95	5.01	5.07	5.13	5.18	5.23
20	4.61	4.70	4.78	4.85	4.92	4.99	5.05	5.10	5.16	5.20
24	4.54	4.63	4.71	4.78	4.85	4.91	4.97	5.02	5.07	5.12
30	4.47	4.56	4.64	4.71	4.77	4.83	4.89	4.94	4.99	5.03
40	4.41	4.49	4.56	4.63	4.69	4.75	4.81	4.86	4.90	4.95
60	4.34	4.42	4.49	4.56	4.62	4.67	4.73	4.78	4.82	4.86
120	4.28	4.35	4.42	4.48	4.54	4.60	4.65	4.69	4.74	4.78
∞	4.21	4.28	4.35	4.41	4.47	4.52	4.57	4.61	4.65	4.69

Table 2. *Upper 5% points of the studentized range*

$\nu \backslash n$	2	3	4	5	6	7	8	9	10
1	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07
2	6.08	8.33	9.80	10.88	11.74	12.44	13.03	13.54	13.99
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92
30	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65
120	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56
∞	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47

$\nu \backslash n$	11	12	13	14	15	16	17	18	19	20
1	50.59	51.96	53.20	54.33	55.36	56.32	57.22	58.04	58.83	59.56
2	14.39	14.75	15.08	15.38	15.65	15.91	16.14	16.37	16.57	16.77
3	9.72	9.95	10.15	10.35	10.52	10.69	10.84	10.98	11.11	11.24
4	8.03	8.21	8.37	8.52	8.66	8.79	8.91	9.03	9.13	9.23
5	7.17	7.32	7.47	7.60	7.72	7.83	7.93	8.03	8.12	8.21
6	6.65	6.79	6.92	7.03	7.14	7.24	7.34	7.43	7.51	7.59
7	6.30	6.43	6.55	6.66	6.76	6.85	6.94	7.02	7.10	7.17
8	6.05	6.18	6.29	6.39	6.48	6.57	6.65	6.73	6.80	6.87
9	5.87	5.98	6.09	6.19	6.28	6.36	6.44	6.51	6.58	6.64
10	5.72	5.83	5.93	6.03	6.11	6.19	6.27	6.34	6.40	6.47
11	5.61	5.71	5.81	5.90	5.98	6.06	6.13	6.20	6.27	6.33
12	5.51	5.61	5.71	5.80	5.88	5.95	6.02	6.09	6.15	6.21
13	5.43	5.53	5.63	5.71	5.79	5.86	5.93	5.99	6.05	6.11
14	5.36	5.46	5.55	5.64	5.71	5.79	5.85	5.91	5.97	6.03
15	5.31	5.40	5.49	5.57	5.65	5.72	5.78	5.85	5.90	5.96
16	5.26	5.35	5.44	5.52	5.59	5.66	5.73	5.79	5.84	5.90
17	5.21	5.31	5.39	5.47	5.54	5.61	5.67	5.73	5.79	5.84
18	5.17	5.27	5.35	5.43	5.50	5.57	5.63	5.69	5.74	5.79
19	5.14	5.23	5.31	5.39	5.46	5.53	5.59	5.65	5.70	5.75
20	5.11	5.20	5.28	5.36	5.43	5.49	5.55	5.61	5.66	5.71
24	5.01	5.10	5.18	5.25	5.32	5.38	5.44	5.49	5.55	5.59
30	4.92	5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.47
40	4.82	4.90	4.98	5.04	5.11	5.16	5.22	5.27	5.31	5.36
60	4.73	4.81	4.88	4.94	5.00	5.06	5.11	5.15	5.20	5.24
120	4.64	4.71	4.78	4.84	4.90	4.95	5.00	5.04	5.09	5.13
∞	4.55	4.62	4.68	4.74	4.80	4.85	4.89	4.93	4.97	5.01

Table 3. *Upper 1 % points of the studentized range*

$\nu \backslash n$	2	3	4	5	6	7	8	9	10
1	90.03	135.0	164.3	185.6	202.2	215.8	227.2	237.0	245.6
2	14.04	19.02	22.29	24.72	26.63	28.20	29.53	30.68	31.69
3	8.26	10.62	12.17	13.33	14.24	15.00	15.64	16.20	16.69
4	6.51	8.12	9.17	9.96	10.58	11.10	11.55	11.93	12.27
5	5.70	6.98	7.80	8.42	8.91	9.32	9.67	9.97	10.24
6	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10
7	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37
8	4.75	5.64	6.20	6.62	6.96	7.24	7.47	7.68	7.86
9	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.33	7.49
10	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21
11	4.39	5.15	5.62	5.97	6.25	6.48	6.67	6.84	6.99
12	4.32	5.05	5.50	5.84	6.10	6.32	6.51	6.67	6.81
13	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67
14	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54
15	4.17	4.84	5.25	5.56	5.80	5.99	6.16	6.31	6.44
16	4.13	4.79	5.19	5.49	5.72	5.92	6.08	6.22	6.35
17	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27
18	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20
19	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14
20	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09
24	3.96	4.55	4.91	5.17	5.37	5.54	5.69	5.81	5.92
30	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76
40	3.82	4.37	4.70	4.93	5.11	5.26	5.39	5.50	5.60
60	3.76	4.28	4.59	4.82	4.99	5.13	5.25	5.36	5.45
120	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30
∞	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16

$\nu \backslash n$	11	12	13	14	15	16	17	18	19	20
1	253.2	260.0	266.2	271.8	277.0	281.8	286.3	290.4	294.3	298.0
2	32.59	33.40	34.13	34.81	35.43	36.00	36.53	37.03	37.50	37.95
3	17.13	17.53	17.89	18.22	18.52	18.81	19.07	19.32	19.55	19.77
4	12.57	12.84	13.09	13.32	13.53	13.73	13.91	14.08	14.24	14.40
5	10.48	10.70	10.89	11.08	11.24	11.40	11.55	11.68	11.81	11.93
6	9.30	9.48	9.65	9.81	9.95	10.08	10.21	10.32	10.43	10.54
7	8.55	8.71	8.86	9.00	9.12	9.24	9.35	9.46	9.55	9.65
8	8.03	8.18	8.31	8.44	8.55	8.66	8.76	8.85	8.94	9.03
9	7.65	7.78	7.91	8.03	8.13	8.23	8.33	8.41	8.49	8.57
10	7.36	7.49	7.60	7.71	7.81	7.91	7.99	8.08	8.15	8.23
11	7.13	7.25	7.36	7.46	7.56	7.65	7.73	7.81	7.88	7.95
12	6.94	7.06	7.17	7.26	7.36	7.44	7.52	7.59	7.66	7.73
13	6.79	6.90	7.01	7.10	7.19	7.27	7.35	7.42	7.48	7.55
14	6.66	6.77	6.87	6.96	7.05	7.13	7.20	7.27	7.33	7.39
15	6.55	6.66	6.76	6.84	6.93	7.00	7.07	7.14	7.20	7.26
16	6.46	6.56	6.66	6.74	6.82	6.90	6.97	7.03	7.09	7.15
17	6.38	6.48	6.57	6.66	6.73	6.81	6.87	6.94	7.00	7.05
18	6.31	6.41	6.50	6.58	6.65	6.73	6.79	6.85	6.91	6.97
19	6.25	6.34	6.43	6.51	6.58	6.65	6.72	6.78	6.84	6.89
20	6.19	6.28	6.37	6.45	6.52	6.59	6.65	6.71	6.77	6.82
24	6.02	6.11	6.19	6.26	6.33	6.39	6.45	6.51	6.56	6.61
30	5.85	5.93	6.01	6.08	6.14	6.20	6.26	6.31	6.36	6.41
40	5.69	5.76	5.83	5.90	5.96	6.02	6.07	6.12	6.16	6.21
60	5.53	5.60	5.67	5.73	5.78	5.84	5.89	5.93	5.97	6.01
120	5.37	5.44	5.50	5.56	5.61	5.66	5.71	5.75	5.79	5.83
∞	5.23	5.29	5.35	5.40	5.45	5.49	5.54	5.57	5.61	5.65

ON THE DISTRIBUTION OF THE EXTREME STUDENTIZED DEVIATE FROM THE SAMPLE MEAN

BY K. C. S. PILLAI* AND BENJAMIN P. TIENZO†

*The Statistical Center, University of the Philippines and the Philippine
Statistical Survey of Households, National Economic Council*

1. INTRODUCTION AND SUMMARY

The distribution of the extreme deviate from the mean of a sample taken from a normal population and that of the studentized extreme deviate from the sample mean are studied and percentage points are obtained for the latter distribution. This paper deals only with small sample sizes.

Let $x_1 \leq x_2 \leq \dots \leq x_n$ be an ordered sample of size n from a normal population, $\bar{x}$ the mean of the sample and s_v the square root of an independent mean square estimate of σ^2 based on ν degrees of freedom, where σ^2 is the square of the standard deviation of the normal population. Then, the distributions considered are those of:

$$u = (x_n - \bar{x})/\sigma \quad \text{or} \quad (\bar{x} - x_1)/\sigma \quad (1)$$

$$t_n = (x_n - \bar{x})/s_v \quad \text{or} \quad (\bar{x} - x_1)/s_v. \quad (2)$$

McKay (1935) derived the distribution of u and indicated its usefulness in testing outlying observations. Nair (1948) made further studies of the distribution and tabulated the probability integral of u for sample sizes varying from 3 to 9. Grubbs (1950) also studied the same distribution problem independently and obtained tables of the probability integral of u for sample sizes ranging from 2 to 25.

The distribution of u as given by McKay (1935) follows the recurrence formula

$$f_n(u) = \frac{n}{\sqrt{(2\pi)}} \exp \left[-\frac{nu^2}{2(n-1)} \right] \sqrt{\left(\frac{n}{n-1} \right)} F_{n-1} \left(\frac{nu}{n-1} \right), \quad (3)$$

where, following Nair (1948), $F_n(u)$, the cumulative distribution function of u , can be given in the form

$$F_n(u) = \frac{\sqrt{n}}{[\sqrt{(2\pi)}]^{n-1}} G_{n-1}(nu)$$

and where

$$G_r(x) = \int_0^x \exp \left[-\frac{t^2}{2r(r+1)} \right] G_{r-1}(t) dt \quad (G_0(x) = 1).$$

Nair (1948), following Hartley's expansion (1944) for what he termed 'studentized integrals', obtained percentage points of the distribution of t_n for sample sizes $n = 3(1)9$ and degrees of freedom $\nu = 10(1)20, 24, 30, 40, 60, 120, \infty$. David (1956) showed that the studentization procedure employed by Nair tended to become unsatisfactory for $\nu < 20$. He provided amended figures, using the same values of ν and adding results for $n = 10, 12$.

Tabulation for $\nu < 10$ was not, however, attempted by previous authors and we shall here make a study of the distribution of u_n and also t_n for $n = 3, 4, 5$, and $\nu \leq 10$. The

* United Nations Senior Adviser in Mathematical Statistics and Visiting Professor of Statistics, Statistical Center, University of the Philippines. Now with United Nations, New York.

† Now Senior Statistician, Bureau of the Census and Statistics, Manila.

percentage points for $\nu < 10$ are particularly useful in many situations, for example, in experiments with small numbers of treatments or small numbers of blocks. The following sections present the study of the distribution problem and an application to testing outlying observations. The percentage points derived from the formulae given in § 4 have been incorporated in the somewhat more extensive tables covering sample sizes up to $n = 12$ given below in the paper by Pillai (1959, p. 473).

2. DEVELOPMENT OF CERTAIN SERIES EXPANSIONS

Pillai (1950) has developed an expansion of the normal probability integral in the form

$$\int_0^w e^{-\frac{1}{2}t^2} dt = w e^{-\frac{1}{2}w^2} (1 + a_2 w^4 + \dots + a_i w^{2i} + \dots), \quad (4)$$

where a_i follows the recurrence relation

$$3(2i+1)a_i - a_{i-1} = (-1)^i / [3^{i-1} \Gamma(i+1)]. \quad (5)$$

The series (4) has been shown to be absolutely convergent by Pillai & Ramachandran (1954).

Now substituting for w , $(3/\sqrt{2})x$ in (4) we get an expansion of the integral $\int_0^{3x/\sqrt{2}} e^{-\frac{1}{2}t^2} dt$. If we denote the resulting series by

$$\int_0^{3x/\sqrt{2}} e^{-\frac{1}{2}t^2} dt = \frac{3x}{\sqrt{2}} e^{-\frac{3}{2}x^2} (1 + b_2 x^4 + \dots + b_i x^{2i} + \dots), \quad (6)$$

then

$$(2i+1)b_i = \frac{3}{2}b_{i-1} + (-\frac{3}{2})^i / i!. \quad (7)$$

If i is large it has been shown by Pillai & Ramachandran (1954) that

$$-a_i/a_{i-1} < 11/30i \quad \text{and hence} \quad -b_i/b_{i-1} < 99/60i,$$

which shows that series (6) is absolutely convergent. The values of the first few b coefficients are given below:

$$\begin{aligned} b_2 &= 0.225, & b_5 &= -0.003,287,337,7, \\ b_3 &= -0.032,142,857, & b_6 &= 0.000,837,638,92, \\ b_4 &= 0.018,080,357, \end{aligned}$$

3. THE DISTRIBUTION OF THE EXTREME DEVIATE FROM THE SAMPLE MEAN

In this section the distribution of the extreme deviate from the mean of a sample from a normal population is developed in series form for $n = 3, 4$ and 5 .

(i) $n = 3$. From (3), $f_3(u)$ can be obtained by putting $n = 3$ and is given by

$$f_3(u) = \frac{3\sqrt{3}}{\pi} e^{-\frac{3}{2}u^2} \int_0^{\frac{3}{2}u} e^{-z^2} dz. \quad (8)$$

Transforming $z = t/\sqrt{2}$ and using the series (6) in (8) we get

$$f_3(u) = \frac{9\sqrt{3}}{2\pi} e^{-\frac{3}{2}u^2} (u + b_2 u^5 + \dots + b_i u^{2i+1} + \dots). \quad (9)$$

In $f_3(u)$, if we retain terms up to and including $b_5 u^{11}$ and approximate the remaining terms in an exponent, we get

$$f_3(u) = \frac{9\sqrt{3}}{2\pi} e^{-\frac{3}{2}u^2} (u + b_2 u^5 + b_3 u^7 + b_4 u^9 + b_5 u^{11} e^{b_6 u^2 b_5}). \quad (10)$$

The c.d.f. of u can be obtained from (10) as a series of incomplete gamma functions, and is given by

$$F_3(U) = \frac{9\sqrt{3}}{2\pi} \left[\sum_{i=0}^4 \frac{1}{2} b_i \left(\frac{2}{3}\right)^{i+1} I_{\frac{3}{2}U^2}(i+1) \Gamma(i+1) + \frac{1}{2} b_5 r^{-6} I_{rU^2}(6) \Gamma(6) \right], \quad (11)$$

where

$$r = \frac{3}{2} - b_6/b_5 \quad \text{and} \quad I_x(p) = \int_0^x e^{-w} w^p dw / \Gamma(p+1).$$

If we make U tend to infinity in (11) we get the total probability. The total probability attained for $F_3(\infty)$ in (11) is 0.999,262. For the 95 and 99 % probability, Grubbs (1950) gives $U = 1.738$ and 2.215 , respectively, which when substituted in (11) give probabilities of 0.950,036 and 0.989,882; hence the expression (11) was considered to give sufficient accuracy for obtaining the upper percentage points.

(ii) $n = 4$. To obtain $f_4(u)$ we substitute $n = 4$ in (3) and use (11). We get

$$f_4(u) = \frac{8}{\sqrt{(6\pi)}} e^{-\frac{2}{3}u^2} F_3\left(\frac{4}{3}u\right). \quad (12)$$

Integrating $f_3(u)$ in (10) by parts in the interval 0 to $\frac{4}{3}u$ and substituting in (12) we arrive at the form

$$f_4(u) = A \{ e^{-\frac{2}{3}u^2} [b + e^{-\frac{2}{3}u^2} (c_0 + c_1 u^2 + \dots + c_4 u^8)] + e^{du^2} (e_0 + e_1 u^2 + \dots + e_5 u^{10}) \}, \quad (13)$$

where $A = 4.571,541,8$, $b = 0.402,768,92$, $d = -3.786,324,8$ and

i	c_i	e_i
0	-0.409,523,81	0.006,754,891,7
1	-0.203,174,60	0.021,072,953
2	-0.270,899,47	0.032,807,204
3	-0.030,099,941	0.034,181,266
4	-0.060,199,882	0.026,658,466
5		0.016,633,060

Integrating (13) in the interval 0 to U we get $F_4(U)$ which could be shown to reduce to the following form involving incomplete gamma functions:

$$F_4(U) = B I_{\frac{2}{3}U^2}(\frac{1}{2}) \Gamma(\frac{1}{2}) + C \sum_{i=1}^5 f_i I_{10U^2/3}(i - \frac{1}{2}) \Gamma(i - \frac{1}{2}) + D \sum_{i=1}^6 g_i I_{dU^2}(i - \frac{1}{2}) \Gamma(i - \frac{1}{2}), \quad (14)$$

where $B = 1.127,546,0$, $C = 1.523,847,3$, $D = -0.002,141,004,0$ and

i	f_i	g_i
1	-0.336,458,14	-3.706,164,5
2	-0.050,077,491	-3.053,611,7
3	-0.020,030,996	-1.257,977,7
4	-0.000,667,699,88	-0.345,494,26
5	-0.000,400,619,93	-0.071,165,578
6		-0.011,727,059

The total probability obtained from $F_4(\infty)$ is 1.000,061, and for the 95 and 99 % points given by Grubbs (1950), we get the values 0.950,062 and 0.990,063.

(iii) $n = 5$. From (3), putting $n = 5$, we get

$$f_5(u) = \frac{5\sqrt{5}}{2\sqrt{(2\pi)}} e^{-\frac{5}{2}u^2} F_4\left(\frac{5}{4}u\right). \quad (15)$$

Integrating $f_4(u)$ in (13) in the interval 0 to $\frac{5}{4}u$, we arrive at $f_5(u)$ in the following form:

$$f_5(u) = E \left\{ e^{-\frac{2.90}{4.8}u^2} [h_0u + h_1u^3 + h_2u^5 + h_3u^7] + e^{au^2} [i_0u + i_1u^3 + i_2u^5 + i_3u^7 + i_4u^9] + F e^{-\frac{5}{2}u^2} \int_0^{\frac{5}{4}u} e^{-\frac{5}{2}w^2} dw + G e^{-\frac{5}{2}u^2} \int_0^{\frac{5}{4}u} e^{-\frac{1}{2}w^2} dw + H e^{-\frac{5}{2}u^2} \int_0^{\frac{5}{4}u} e^{dw^2} dw \right\}, \quad (16)$$

where $E = 10.195,247$, $F = 0.402,768,92$, $G = -0.463,009,52$, $H = 0.013,920,354$, $a = -6.541,132,5$ and

i	h_i	i_i
0	0.066,857,143	
1	0.099,867,725	-0.008,956,828
2	0.042,713,845	-0.021,607,151
3	0.043,058,311	-0.031,069,928
4		-0.029,234,091
		-0.016,364,940

4. THE DISTRIBUTION OF THE EXTREME STUDENTIZED DEVIATE FROM THE SAMPLE MEAN

The distribution of $t_n = (x_n - \bar{x})/s_v$ or $(\bar{x} - x_1)/s_v$ could be obtained by using $f_n(u)$ derived in the previous section and $f(s_v)$ given by

$$f(s_v) = [2(\frac{1}{2}\nu)^{\frac{1}{2}\nu}/\Gamma(\frac{1}{2}\nu)] s_v^{\nu-1} e^{-\frac{1}{2}\nu s_v^2}. \quad (17)$$

Since s_v is independent of u , the simultaneous distribution of u and s_v is given by

$$f(u, s_v) = f_n(u) f(s_v). \quad (18)$$

In (18), if we use the transformation

$$t_n = u/s_v \quad \text{and} \quad s_v = s_v \quad (19)$$

and integrate s_v in the interval 0 to ∞ we get $f(t_n)$. Let us study the distribution of t_3 , t_4 and t_5 separately.

(i) $n = 3$. Multiplying (10) by (17), applying the transformation (19) with $n = 3$ and integrating s_v from 0 to ∞ we get $f(t_3)$. $f(t_3)$ can be shown to be of the form

$$f(t_3) = [18\sqrt{3}(\frac{1}{2}\nu)^{\frac{1}{2}\nu}/\{2\pi\Gamma(\frac{1}{2}\nu)\}] \left\{ \frac{t_3}{2} \left(\frac{2}{3t_3^2 + \nu} \right)^{\frac{1}{2}\nu+1} \Gamma\left(\frac{\nu+2}{2}\right) + \sum_{i=2}^4 \frac{b_i t_3^{2i+1}}{2} \left(\frac{2}{3t_3^2 + \nu} \right)^{\frac{1}{2}\nu+i+1} \Gamma\left(\frac{\nu}{2} + i + 1\right) + \frac{b_5 t_3^{11}}{2} \left[\frac{2b_5}{b_5(3t_3^2 + \nu) - 2b_6 t_3^2} \right]^{\frac{1}{2}(\nu+12)} \Gamma\left(\frac{\nu+12}{2}\right) \right\}. \quad (20)$$

Terms involved in $f(t_3)$ can be reduced to β -functions since t_3 varies from 0 to ∞ . Hence the C.D.F. of t_3 can be given in the form

$$F(t_3) = K \left\{ \sum_{i=0}^4 \frac{b_i \Gamma(\frac{1}{2}\nu + i + 1) 2^{i-1}}{3^{i+1} \Gamma(\frac{1}{2}\nu)} \int_0^\infty \theta^i (1-\theta)^{\frac{1}{2}\nu-1} d\theta + \frac{2^4 b_5 \Gamma(\frac{1}{2}\nu + 6)}{L \Gamma(\frac{1}{2}\nu)} \int_0^\beta \theta^5 (1-\theta)^{\frac{1}{2}\nu-1} d\theta \right\}, \quad (21)$$

where $K = 4.961,960,1$, $L = 1,868.775,6$,
 $\alpha = 3t_3^2/(\nu + 3t_3^2)$, $\beta = 3.509,615,4t_3^2/(\nu + 3.509,615,4t_3^2)$.

In arriving at (21) the transformation used for all terms except the last one in (20) is $\theta = \alpha$, and in the last term of (20) is $\theta = \beta$.

Alternatively, for odd values of ν , it is possible to obtain the distribution of t_3 separately for each ν in a convenient form involving β -functions. The distribution obtained in this manner is exact. For example, when $\nu = 1$, multiplying $f_3(u)$ and $f(s_1)$, using transformation (19) with $\nu = 1$ and integrating with respect to s_1 by parts, we get

$$f_{\nu=1}(t_3) = [(9\sqrt{3})/\pi] [t/(3t^2 + 2)(6t^2 + 1)^{\frac{1}{2}}] \quad (22)$$

and $F_{\nu=1}(t_3) = [(3\sqrt{3})/(8\pi)] \int_0^{6t^2/(1+6t^2)} (1-\theta)^{-\frac{1}{2}} (1-3\theta/4)^{-1} d\theta. \quad (23)$

The other cases also could be attempted in a similar manner.

(ii) $n = 4$. Similarly, using $f_4(u)$ given in (13) and $f(s_\nu)$ in (17), we get

$$f(t_4) = \frac{1}{2}A \left[b(\frac{3}{2})^{\frac{1}{2}} \Gamma(\frac{1}{2}) + \sum_{i=0}^4 c_i (\frac{3}{10})^{\frac{1}{2}(2i+1)} \Gamma\left(\frac{2i+1}{2}\right) + \sum_{i=0}^5 e_i (2/7.572,649,6)^{\frac{1}{2}(2i+1)} \Gamma\left(\frac{2i+1}{2}\right) \right] \quad (24)$$

and
$$F(t_4) = Mb \frac{\Gamma\{\frac{1}{2}(\nu+1)\}}{\Gamma(\frac{1}{2}\nu)} \int_0^{m'} \theta^{-\frac{1}{2}} (1-\theta)^{\frac{1}{2}\nu-1} d\theta$$

$$+ N \sum_{i=0}^4 c_i (0.3)^{\frac{1}{2}(2i+1)} \frac{\Gamma\{\frac{1}{2}(\nu+2i+1)\}}{\Gamma(\frac{1}{2}\nu)} \int_0^{n'} \theta^{\frac{1}{2}(2i-1)} (1-\theta)^{\frac{1}{2}\nu-1} d\theta$$

$$+ N \sum_{i=0}^5 e_i (-d)^{-\frac{1}{2}(2i+1)} \frac{\Gamma\{\frac{1}{2}(\nu+2i+1)\}}{\Gamma(\frac{1}{2}\nu)} \int_0^{p'} \theta^{\frac{1}{2}(2i-1)} (1-\theta)^{\frac{1}{2}\nu-1} d\theta, \quad (25)$$

where $M = 2.799,486,2$, $N = 2.285,770,9$, $m' = 4t_4^2/(3\nu + 4t_4^2)$,
 $n' = 20t_4^2/(3\nu + 20t_4^2)$, $p' = 7.572,649,6t_4^2/(\nu + 7.572,649,6t_4^2)$.

(iii) $n = 5$. The method used for this case is similar to that adopted for $f_{\nu=1}(t_3)$. $f(t_5)$ for $\nu = 1, 3, 5, 7, 9$ and 11 have been evaluated separately as well as the corresponding C.D.F.'s. The expressions for the C.D.F. are given below for the different cases.

$$F(t_5) = \sum_{i=0}^{\frac{1}{2}(\nu-1)} \nu S_i \int_0^{k_\nu} (1-\theta)^{\frac{1}{2}(\nu-2)} (1-5\theta/8)^{-(i+1)} d\theta$$

$$+ \sum_{i=0}^{\frac{1}{2}(\nu-1)} \nu T_i \int_0^{l_\nu} (1-\theta)^{\frac{1}{2}(\nu-2)} (1-25\theta/28)^{-(i+1)} d\theta$$

$$+ \sum_{i=0}^{\frac{1}{2}(\nu-1)} \nu U_i \int_0^{m_\nu} (1-\theta)^{\frac{1}{2}(\nu-2)} (1-0.904,450,80\theta)^{-(i+1)} d\theta$$

$$+ \sum_{i=0}^3 \nu V_i \int_0^{l_\nu} (1-\theta)^{\frac{1}{2}(\nu-2)} \theta^i d\theta + \sum_{i=0}^4 \nu W_i \int_0^{m_\nu} (1-\theta)^{\frac{1}{2}(\nu-2)} \theta^i d\theta \quad (\nu = 1, 3, 5, 7, 9, 11). \quad (26)$$

The values of νS_i , νT_i , νU_i , k_ν , l_ν , m_ν , νV_i and νW_i , are given in Tienzo (1958).

5. PERCENTAGE POINTS FOR THE EXTREME STUDENTIZED DEVIATE FROM THE SAMPLE MEAN

The percentage points for $n = 3, 4$ and 5 were computed from the formulae for $F(t_n)$ given in the previous section. For $n = 5$ and even degrees of freedom, the values were obtained by interpolation. For $n = 2$ the values were computed by dividing by 2 corresponding percentage points of the studentized range (Pillai, 1957). All these values have been incorporated in the more extensive table of 5 and 1% points for sample sizes up to $n = 12$, printed on p. 473.

6. APPLICATION OF THE TEST

Cochran & Cox (§ 4.23, 1950) have given an example of a randomized block experiment carried out by the North Carolina Agricultural Experiment Station in 1944. The experiment tested the effect of five levels of application of potash, supplying respectively 36, 54, 72, 108 and 144 lb. K_2O per acre, on the yield and properties of cotton. The experiment was arranged in three randomized blocks of five plots each. The means for the five levels were respectively 7.85, 8.05, 7.74, 7.51 and 7.45 with general mean 7.72.

Analysis of variance				
Source of variation	D.F.	S.S.	M.S.	F
Replications	2	0.097,1	—	—
Treatment	4	0.732,4	0.183,1	4.19
Error	8	0.349,5	—	(Significant at 95 % level
Total	14	1.179,0	0.043,7	but not significant at
				the 99 % level)
			—	—

For testing the difference of the largest mean from the general mean, we compute $t_5 = 2.73$. Referring to Table 1 on p. 473, we see that this value is significant at the 95 % but not significant at the 99 % level.

Considering the smallest mean, the corresponding value of t_5 is 2.24. Referring to the same Table, it is found not to be significant at the 95 % level.

The authors wish to acknowledge the facilities offered by the Statistical Center, University of the Philippines, in the preparation of this paper.

REFERENCES

COCHRAN, W. G. & COX, G. M. (1950). *Experimental Design*. New York: John Wiley and Sons, Inc.

DAVID, H. A. (1956). *Biometrika*, **43**, 449.

GRUBBS, F. E. (1950). *Ann. Math. Statist.* **21**, 27.

HARTLEY, H. O. (1944). *Biometrika*, **33**, 173.

McKAY, A. T. (1935). *Biometrika*, **27**, 466.

NAIR, K. R. (1948). *Biometrika*, **35**, 118.

PILLAI, K. C. S. (1950). *Ann. Math. Statist.* **21**, 100.

PILLAI, K. C. S. (1957). *Concise Tables for Statisticians*. The Statistical Center, University of the Philippines.

PILLAI, K. C. S. (1959). *Biometrika*, **46**, 473-4.

PILLAI, K. C. S. & RAMACHANDRAN, K. V. (1954). *Ann. Math. Statist.* **25**, 565.

TENZO, B. P. (1958). On the distribution of the extreme studentized deviate from the sample mean. Unpublished Thesis, The Statistical Center, University of the Philippines.

Upper percentage points of the extreme studentized deviate from the sample mean

By K. C. S. PILLAI*

The Statistical Center, University of the Philippines

Let $x_1 \leq x_2 \leq \dots \leq x_n$ be an ordered sample of size n from a normal population, $\bar{x}$, the mean of the sample and s_v , the square root of an independent mean square estimate of σ^2 based on ν degrees of freedom, where σ^2 is the square of the standard deviation of the normal population. Then the extreme studentized deviate from the sample mean is defined as $t_n = (x_n - \bar{x})/s_v$ or $(\bar{x} - x_1)/s_v$. Table 1 gives the upper 5 and 1 % points of t_n for sample sizes ranging from 2 to 12 and degrees of freedom from 1 to 10, thus filling the gap that existed in the tables of upper percentage points of t_n .†

Table 1. Upper percentage points of the extreme studentized deviate from the sample mean
($x_n - \bar{x})/s_v$ or $(\bar{x} - x_1)/s_v$

$\nu \backslash n$	2	3	4	5	6	7	8	9	10	12
5 % points										
1	9.0	13.5	16.4	19	20	22	23	24	25	26
2	3.04	4.23	4.98	5.5	6.0	6.3	6.6	6.9	7.1	7.5
3	2.25	3.03	3.50	3.88	4.15	4.36	4.55	4.72	4.86	5.11
4	1.96	2.58	2.98	3.26	3.48	3.65	3.80	3.93	4.05	4.24
5	1.82	2.37	2.71	2.95	3.15	3.30	3.43	3.54	3.64	3.80
6										
7	1.73	2.24	2.55	2.78	2.95	3.09	3.21	3.31	3.39	3.54
8	1.67	2.15	2.45	2.66	2.82	2.95	3.06	3.15	3.23	3.37
9	1.63	2.09	2.37	2.57	2.72	2.85	2.95	3.04	3.12	3.25
10	1.60	2.04	2.32	2.51	2.65	2.78	2.87	2.96	3.03	3.15
	1.58	2.01	2.27	2.46	2.60	2.72	2.81	2.89	2.96	3.08
1 % points										
1	45.0	68	82	93	101	108	114	119	123	130
2	7.0	9.9	11.3	12.6	13.6	14.4	15.0	15.6	16.1	16.9
3	4.13	5.5	6.3	6.9	7.3	7.7	8.1	8.4	8.6	9.0
4	3.26	4.23	4.81	5.23	5.54	5.80	6.03	6.22	6.39	6.68
5	2.85	3.65	4.11	4.45	4.70	4.93	5.11	5.26	5.39	5.62
6										
7	2.62	3.32	3.72	4.02	4.24	4.43	4.58	4.71	4.82	5.01
8	2.48	3.11	3.48	3.74	3.94	4.11	4.25	4.37	4.46	4.63
9	2.37	2.96	3.31	3.56	3.74	3.89	4.02	4.13	4.22	4.38
10	2.30	2.86	3.19	3.41	3.59	3.73	3.86	3.95	4.04	4.19
	2.24	2.78	3.10	3.32	3.48	3.62	3.73	3.82	3.90	4.0

* United Nations Senior Adviser in Mathematical Statistics and Visiting Professor of Statistics now with the United Nations, New York.
† The columns for $n = 2$ and the rows for $\nu = 10$ are not new but are included for the convenience of the table user.

Table 1 was prepared in three steps. First, a table of trial values was obtained by taking one-half of the corresponding percentage points of the studentized range, $q = (x_n - x_1)/s_\nu$, from Pillai's Tables (1957), and adjusting these values after noting the differences between them and the percentage points of t_n given by David (1956) and Pillai & Tienzo (1959). Secondly, based on the table of trial values and employing David's method (1956) of solving the equation

$$\int_0^\infty f(u_n) \int_0^{u_n/R_\alpha} f(s_\nu) ds_\nu du_n = \alpha, \quad (1)$$

(where R_α is the upper 100α percentage point of the extreme studentized deviate and u_n the extreme deviate from the sample mean) percentage points of t_n were obtained for $\nu = 1, 2, 4, 6$ and 8 and values of n starting from 6 . Since the integral for s_ν is exactly evaluable for even degrees of freedom, equation (1) was solved for R_α by numerical integration using Grubbs' tables (1950) of the cumulative distribution function of u_n . For $\nu = 1$ both Grubbs' tables and normal probability tables were used. Lastly, percentage points for $\nu = 3, 5, 7$ and 9 and n ranging from 6 to 12 were obtained through adjustment of the trial values by the average of the differences observed for the two neighbouring values from the corresponding trial values for even degrees of freedom and the same value of n .

I am grateful to Professor E. S. Pearson for his suggestion of filling the gap between the table of David and that of Pillai & Tienzo.

My thanks are due to Miss Aurora Abesamis for the computational assistance in preparing Table 1. I also wish to acknowledge the facilities offered by the Statistical Center in the preparation of this paper.

REFERENCES

- DAVID, H. A. (1956). *Biometrika*, **43**, 449.
 GRUBBS, F. E. (1950). *Ann. Math. Stat.* **21**, 27.
 PILLAI, K. C. S. (1957). *Concise Tables for Statisticians*. The Statistical Center, University of the Philippines.
 PILLAI, K. C. S. & TIENZO, B. P. (1959). *Biometrika*, **46**, 467-72.

MISCELLANEA

The asymptotic efficiency of the χ_r^2 -test for a balanced incomplete block design*

BY PH. VAN ELTEREN AND G. E. NOETHER†

Mathematical Centre, Amsterdam, and Boston University

Friedman (1937) has shown how n treatments can be compared on the basis of the rankings of m 'observers'. This method has been extended by Durbin (1951) to cover a balanced incomplete block design, i.e. the case when each observer ranks $k \leq n$ treatments exactly once and each treatment is compared with any other treatment exactly λ times. We want to find the asymptotic ($m \rightarrow \infty$) relative efficiency (in the sense of Pitman, see e.g. Hannan, 1956) of Durbin's test with respect to the usual analysis of variance test for a balanced incomplete block design.

We note for future reference that

$$\lambda = \frac{l(k-1)}{n-1}, \quad (1)$$

where l is the total number of times a given treatment is used (replications).

The asymptotic relative efficiency is most easily obtained with the help of a formula given explicitly for the first time by Hannan (1956). It may be stated roughly in the following way. If both test statistics have, under the alternative hypothesis, non-central χ -square distributions with the same number of degrees of freedom, the asymptotic relative efficiency of one test with respect to the other test is equal to the ratio of the two non-centrality factors after the alternatives have been set equal.

Essentially, then, all we have to do is to compute the two non-centrality factors which we shall denote by d_r^2 and d_F^2 for the rank and F -tests, respectively. The conditions for the applicability of Hannan's formula can be shown to hold provided the underlying distributions satisfy some very general regularity conditions. These details will not be presented in this paper, since similar considerations have been given in several other papers, e.g. Andrews (1954), Benard & van Elteren (1953) and Bradley (1955).

Before computing the non-centrality factors, we have to specify the mathematical model which we are going to consider. The usual analysis of variance model suggests the following approach.

Let $F(x)$ be a continuous cumulative distribution function with density function $f(x) = F'(x)$. Let $\mathbf{x}_{\mu\nu}$ ($\mu = 1, 2, \dots, m$; $\nu = 1, 2, \dots, n$)‡ be the chance variable associated with the observation of the μ th observer (block, in the analysis of variance) on the ν th treatment. It is then assumed that the distribution $F_{\mu\nu}(x)$ of $\mathbf{x}_{\mu\nu}$ is given by

$$F_{\mu\nu}(x) = F(x + \theta_\nu + \eta_\mu),$$

where, without loss of generality, we may assume that $\sum_\nu \theta_\nu = 0$. The null hypothesis to be tested is that $\theta_1 = \theta_2 = \dots = \theta_n = 0$. The alternatives to be considered specify that for a given number of replications l ,

$$\theta_\nu = \theta_{\nu l} = \frac{\delta_\nu}{\sqrt{l}},$$

where the δ_ν are given constants satisfying

$$\sum_\nu \delta_\nu = 0. \quad (2)$$

For this model, we find (e.g. Anderson & Bancroft, 1952, §19.3) that

$$d_F^2 = lC \sum_\nu \theta_\nu^2 / \sigma^2 = C \sum_\nu \delta_\nu^2 / \sigma^2, \quad (3)$$

where σ^2 is the variance associated with $F(x)$ and

$$C = \frac{n(k-1)}{k(n-1)} \quad (4)$$

is the efficiency factor of the given incomplete block design. The fact that for large m , F has approximately a non-central χ -square distribution follows from an argument similar to that given in Andrews (1954). For computing d_r^2 , let us introduce the following notation. Let $r_{\mu\nu}$ stand for the rank which the μ th observer assigns to the ν th treatment (assuming that he considers the ν th treatment), and let $R_\nu = \sum_\mu^{(v)} r_{\mu\nu}$

* Report SP 63 of the Mathematical Centre, Amsterdam.

† Work supported by the Office of Naval Research.

‡ Here and below we use bold type to indicate chance variables.

where $\sum^{(v)}$ denotes summation over those l observers who have rated treatment v . If we then set $u_v = R_v - \frac{1}{2}l(k+1)$, Durbin's test statistic is, except for a constant factor given by

$$\chi_r^2 = \frac{12}{n\lambda(k+1)} \sum_{v=1}^n u_v^2,$$

which, for large l , has approximately a non-central χ -square distribution with $(n-1)$ degrees of freedom. We then have

$$d_r^2 = \frac{12}{n\lambda(k+1)} \sum_v \{\mathcal{E}(u_v)\}^2.$$

Now

$$\mathcal{E}(R_i) = \sum_{\mu}^{(i)} \sum_{j(\mu) \neq i} P\{x_{\mu i} > x_{\mu j(\mu)}\} + l, \quad (5)$$

where $j(\mu)$ refers to the treatments rated by the μ th observer. Since

$$\begin{aligned} P\{x_{\mu i} > x_{\mu j(\mu)}\} &= \int F(x + \theta_{j(\mu)} + \eta_{\mu}) f(x + \theta_i + \eta_{\mu}) dx \\ &= \int F(x + \theta_{j(\mu)} - \theta_i) f(x) dx = \int F\left(x + \frac{\delta_{j(\mu)} - \delta_i}{\sqrt{l}}\right) f(x) dx \sim \frac{1}{2} + \frac{\delta_{j(\mu)} - \delta_i}{\sqrt{l}} \int f^2(x) dx, \end{aligned}$$

and since every treatment occurs exactly λ times together with every other treatment, (5), in view of (1) and (2), becomes

$$\mathcal{E}(R_i) \sim \lambda \sum_{\substack{j=1 \\ j \neq i}}^n \left[\frac{1}{2} + \frac{\delta_j - \delta_i}{\sqrt{l}} \int f^2(x) dx \right] + l = \frac{l(k+1)}{2} - \frac{n\lambda}{\sqrt{l}} \delta_i \int f^2(x) dx.$$

Thus

$$\mathcal{E}(u_v) \sim -\frac{n\lambda}{\sqrt{l}} \delta_i \int f^2(x) dx,$$

and, finally,

$$d_r^2 \sim \frac{12n(k-1)}{(k+1)(n-1)} \left[\int f^2(x) dx \right]^2 \sum_v \delta_v^2. \quad (6)$$

From (3), (4) and (6), the asymptotic relative efficiency E_k of the χ_r^2 -test with respect to the F -test is found to be

$$E_k = \frac{12k}{k+1} \left[\sigma \int f^2(x) dx \right]^2. \quad (7)$$

It is interesting to note that (7) depends only on the block size k , and not on the number of treatments. We obtain the relative efficiency of the Friedman test by replacing k by n in (7). In the special case when $f(x)$ is normal, (7) becomes

$$E_k(N) = \frac{3k}{\pi(k+1)}, \quad (7N)$$

some values of which are tabulated below:

k	2	3	4	5	6	7	8	9	10	17	∞
$E_k(N)$	0.64	0.72	0.76	0.80	0.82	0.84	0.85	0.86	0.87	> 0.90	0.95

When $k=2$, we have the case of paired comparisons where each observer is asked to compare only two treatments. In this case, (7) becomes

$$E_2 = 8 \left[\sigma \int f^2(x) dx \right]^2, \quad (8)$$

and (7N),

$$E_2(N) = \frac{2}{\pi} = 0.637. \quad (8N)$$

The problem of paired comparisons has been of considerable interest during the last few years and many methods for treating the problem have been offered. It is easy to show that the χ_r^2 -test for $k=2$ is equal to the asymptotic form of the Bradley-Terry test discussed by Bradley (1955).

Bradley finds $n/\{\pi(n-1)\}$ as the asymptotic efficiency of his T -test relative to the analysis of variance test in case of normality where, following our notation, n is the number of treatments involved. The difference between Bradley's result and our own result (8N) is due to the fact that Bradley assumes a one way classification for the analysis of variance with l observations in each class instead of the balanced incomplete block design which we have considered. Since the variance σ^2 in (8) clearly refers to the within block variability, the latter model seems to be more appropriate.

If we multiply Bradley's result by $1/C$, where C is given by (4) with $k=2$, lC being the effective number of replications of the balanced incomplete block design, we get (8N).

As far as the χ_r^2 -test is concerned, instead of the model

$$F_{\mu\nu}(x) = F(x + \theta_\nu + \eta_\mu),$$

we could have considered the more general model

$$F_{\mu\nu}(x) = F_{\mu}(x + \theta_{\nu}),$$

where, for different μ , the $F_{\mu}(x)$ may be different cumulative distribution functions. Nothing much is gained as long as we are interested in an incomplete block design, since the non-centrality parameter d_r^2 , and therefore the power of the χ_r^2 -test, will depend on which particular k treatments are put in which particular block. However, in the case of Friedman's test when $k = n$, easy computations show that

$$d_r^2 = \frac{12n}{m^2(n+1)} \left[\sum_{\mu=1}^m \int f_{\mu}^2(x) dx \right]^2 \sum_{\nu=1}^n \delta_{\nu}^2, \quad (9)$$

where $f_{\mu}(x) = F'_{\mu}(x)$. (9) can be used to compute the asymptotic power of Friedman's test for this more general alternative.

For $k = n = 2$, the χ_r^2 -test is equivalent to the two-sided sign test, a test whose properties have been more thoroughly investigated than those of any other distribution-free test. (8) then becomes the asymptotic efficiency of the sign test relative to the t -test, a quantity which is usually given (e.g. Hodges & Lehmann, 1956) as

$$E_{s,t} = 4\sigma^2 f^2(0). \quad (10)$$

The explanation is that (10) refers to the one-sample sign test while (8) refers to the two-sample sign test, a distinction which is rarely made. (10) can be used for the two-sample case if f is interpreted as the density of $\mathbf{x} - \mathbf{y}$, where $\mathbf{x}$ and $\mathbf{y}$ are the two chance variables under consideration. From a practical point of view, it is more important to know the efficiency in terms of the individual distribution of $\mathbf{x}$ and $\mathbf{y}$. Then the answer is given by (8). As far as the two-sample sign test is concerned, it is, of course, easy to obtain (8) directly without reference to the χ_r^2 -test.

The results obtained in this paper are asymptotic, i.e. valid for large numbers of replications. Nothing seems to be known about the relative efficiency of the χ_r^2 -test for small numbers of replications except in the case $k = n = 2$. Since in this particular case the relative efficiency for a small number of replications is known to be much higher (in the case of normality) than indicated by the asymptotic formula, it seems reasonable to assume that the asymptotic values derived in this paper can be considered minimum values for the corresponding efficiencies.

REFERENCES

- ANDERSON, R. L. & BANCROFT, T. A. (1952). *Statistical Theory in Research*. New York: McGraw-Hill.
- ANDREWS, F. C. (1954). Asymptotic behaviour of some rank tests for analysis of variance. *Ann. Math. Statist.* **25**, 724-35.
- BENARD, A. & VAN ELTEREN, PH. (1953). A generalization of the method of m rankings. *Proc. Kon. Ned. Ak. van Wet. A*, **56**. *Indag. math.* **15**, 358-69.
- BRADLEY, R. A. (1955). Rank analysis of incomplete block designs, III. *Biometrika*, **42**, 450-70.
- DURBIN, J. (1951). Incomplete blocks in ranking experiments. *Brit. J. Psychol.* **4**, 85-90.
- FRIEDMAN, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. Statist. Ass.* **32**, 675-99.
- HANNAN, E. J. (1956). The asymptotic powers of certain tests based on multiple correlations. *J. R. Statist. Soc. B*, **18**, 227-33.
- HODGES, J. L., JR. & LEHMANN, E. L. (1956). The efficiency of some non-parametric competitors of the t -test. *Ann. Math. Statist.* **27**, 324-35.

A note on the application of Quenouille's method of bias reduction to the estimation of ratios

By J. DURBIN

Research Techniques Division, London School of Economics

1. An ingenious device for reducing estimation bias from $O(n^{-1})$ to $O(n^{-2})$ has been proposed by Quenouille (1949) and has been considered further in a recent paper by the same author (Quenouille, 1956). The simplest form of the device is as follows. Suppose we have an estimator t_1 based on n observations whose bias is $cn^{-1} + O(n^{-2})$, where c is a constant. Let t_2 and t_3 denote the estimators calculated from the two samples of $\frac{1}{2}n$ observations obtained by dividing the original sample into two equal parts. Then the estimator $t = 2t_1 - \frac{1}{2}(t_2 + t_3)$ has bias of order n^{-2} at most.

It might, perhaps, have been expected that this reduction in bias could only be achieved at the expense of a corresponding increase of variance. However, in his 1956 paper Quenouille showed that any such increase in variance is of small order in n compared with the variance itself. The purpose of the present note is to demonstrate the rather surprising result that for an important class of estimators Quenouille's device actually reduces the variance. The estimators concerned are ratio estimators of the form $r = y/x$ where the regression of y on x is linear and where x is normally distributed. We use asymptotic expansions taking terms in n^{-1} as far as n^{-3} . Later on we consider other ratio estimators for which reduction in bias is such that the mean-square error is reduced. This is an exact result for any sample size. The implication of these findings is that Quenouille's device should frequently lead to an improvement in the accuracy of estimation, particularly in sample surveys where ratio estimators are very

2. Suppose x is a normal variable with variance $O(n^{-1})$. Choose the units of measurement so that $E(x) = 1$, and let $x = 1 - \xi$. Denote $V(x)$ by h . For sufficiently large n we have

$$E(x^{-1}) = E(1 + \xi + \xi^2 + \dots).$$

Taking the first four non-vanishing terms we find

$$E(x^{-1}) = 1 + h + 3h^2 + 15h^3 + O(n^{-4}). \quad (1)$$

Similarly,

$$\begin{aligned} E(x^{-2}) &= E(1 + 2\xi + 3\xi^2 + \dots) \\ &= 1 + 3h + 15h^2 + 105h^3 + O(n^{-4}). \end{aligned} \quad (2)$$

From now on we neglect the terms of $O(n^{-4})$.

Suppose that $r = y/x$ is a ratio estimator of $\rho = E(y)/E(x)$, where the regression of y on x is linear of the form $y = \alpha + \beta x + u$, and where the variance of u is a constant δ of $O(n^{-1})$. Then $\rho = \alpha + \beta$. Also $E(r) = \alpha E(x^{-1}) + \beta$. Using (1) the bias is therefore

$$\alpha(h + 3h^2 + 15h^3), \quad (3)$$

which is $O(n^{-1})$.

Now $r = \beta + (\alpha + u)/x$. Consequently

$$\begin{aligned} E(r - \beta)^2 &= (\alpha^2 + \delta) E(x^{-2}) \\ &= (\alpha^2 + \delta) (1 + 3h + 15h^2 + 105h^3) \end{aligned} \quad (4)$$

using (2). From (3) and (4) we find

$$V(r) = \alpha^2(h + 8h^2 + 69h^3) + \delta(1 + 3h + 15h^2 + 105h^3). \quad (5)$$

3. On splitting the sample into two halves we have the ratio estimators $r_1 = y_1/x_1$ and $r_2 = y_2/x_2$, where $y = \frac{1}{2}(y_1 + y_2)$ and $x = \frac{1}{2}(x_1 + x_2)$. Suppose that $y_1 = \alpha + \beta x_1 + u_1$ and $y_2 = \alpha + \beta x_2 + u_2$ where $E(u_i | x_i) = 0$ and $E(u_i^2 | x_i) = 2\delta$ ($i = 1, 2$). Then $u = \frac{1}{2}(u_1 + u_2)$. Quenouille's estimator is

$$\begin{aligned} t &= 2r - \frac{1}{2}(r_1 + r_2) \\ &= \beta + \alpha \left\{ \frac{2}{x} - \frac{1}{2} \left(\frac{1}{x_1} + \frac{1}{x_2} \right) \right\} + \frac{u_1 + u_2}{x} - \frac{1}{2} \left(\frac{u_1}{x_1} + \frac{u_2}{x_2} \right). \end{aligned}$$

Since $V(x_1) = V(x_2) = 2h$ we find on replacing h by $2h$ in (1) and (2) $E(x_i^{-1}) = 1 + 2h + 12h^2 + 120h^3$ and $E(x_i^{-2}) = 1 + 6h + 60h^2 + 840h^3$ ($i = 1, 2$). Thus

$$\begin{aligned} E(t - \beta) &= \alpha E \left\{ \frac{2}{x} - \frac{1}{2} \left(\frac{1}{x_1} + \frac{1}{x_2} \right) \right\} \\ &= \alpha(1 - 6h^2 - 90h^3). \end{aligned} \quad (6)$$

Thus the bias is $\alpha(6h^2 + 90h^3)$, which is $O(n^{-2})$.

Now

$$\begin{aligned} E \left\{ \frac{2}{x} - \frac{1}{2} \left(\frac{1}{x_1} + \frac{1}{x_2} \right) \right\}^2 &= E \left\{ \frac{4}{x^2} - \frac{2}{x} \left(\frac{1}{x_1} + \frac{1}{x_2} \right) + \frac{1}{4} \left(\frac{1}{x_1^2} + \frac{1}{x_2^2} + \frac{2}{x_1 x_2} \right) \right\} \\ &= E \left\{ \frac{4}{x^2} + \frac{1}{4} \left(\frac{1}{x_1^2} + \frac{1}{x_2^2} \right) - \frac{7}{2} \left(\frac{1}{x_1 x_2} \right) \right\} \\ &= 4(1 + 3h + 15h^2 + 105h^3) + \frac{1}{2}(1 + 6h + 60h^2 + 840h^3) - \frac{7}{2}(1 + 2h + 12h^2 + 120h^3)^2 \\ &= 1 + h - 8h^2 - 168h^3. \end{aligned}$$

Also,

$$E \left[\left(\frac{u_1 + u_2}{x} - \frac{1}{2} \left(\frac{u_1}{x_1} + \frac{u_2}{x_2} \right) \right)^2 \middle| x_1, x_2 \right] = 2\delta \left[\left(\frac{1}{x} - \frac{1}{2x_1} \right)^2 + \left(\frac{1}{x} - \frac{1}{2x_2} \right)^2 \right],$$

whence

$$E \left(\frac{u_1 + u_2}{x} - \frac{1}{2} \left(\frac{u_1}{x_1} + \frac{u_2}{x_2} \right) \right)^2 = 2\delta E \left\{ \frac{2}{x^2} + \frac{1}{4} \left(\frac{1}{x_1^2} + \frac{1}{x_2^2} \right) - \frac{2}{x_1 x_2} \right\} \\ = 2\delta \left(\frac{1}{2} + h + 4h^2 + 54h^3 \right).$$

Consequently,

$$E(t - \beta)^2 = \alpha^2(1 + h - 8h^2 - 168h^3) + \delta(1 + 2h + 8h^2 + 108h^3). \quad (7)$$

From (6) and (7) we find

$$V(t) = V(t - \beta) = \alpha^2(h + 4h^2 + 12h^3) + \delta(1 + 2h + 8h^2 + 108h^3). \quad (8)$$

Comparing (5) and (8) we see that in spite of the fact that, for sufficiently large n , t has a smaller bias than r , $V(t)$ is smaller than $V(r)$.

4. The error entailed by using the first few terms of the expansion of $(1 - \xi)^{-1}$ in place of $1/x$ is $1/x - (1 + \xi + \dots + \xi^7) = \xi^8/x$. The expected value of this is approximately $105h^4$. The error in the corresponding expansion for $1/x^2$ is $1/x^2 - (1 + 2\xi + \dots + 8\xi^7)$ which has an expectation of approximately $945h^4$. Now h is the square of the coefficient of variation of x . Even for a coefficient of variation as large as 20% we have $945h^4 < \frac{1}{400}$. Thus one would expect the errors arising from the use of the expansions to be small even for quite moderate values of n .

5. Let us now consider the case where the regression of y on x has the same form as has been stipulated above, but where x now has the Γ -distribution with density $x^{m-1} \exp(-x)/\Gamma(m)$. It follows easily that

$$E\left(\frac{1}{x}\right) = \frac{1}{m-1}, \quad E\left(\frac{1}{x^2}\right) = \frac{1}{(m-1)(m-2)}.$$

Taking $x = \frac{1}{2}(x_1 + x_2)$ as before, and supposing that $\frac{1}{2}x_1$ and $\frac{1}{2}x_2$ are independent Gamma variables with parameters $\frac{1}{2}m$, we have

$$E\left(\frac{1}{x_i}\right) = \frac{1}{m-2}, \quad E\left(\frac{1}{x_i^2}\right) = \frac{1}{(m-2)(m-4)} \quad (i = 1, 2).$$

Thus we have

$$E(r - \beta) = \frac{\alpha}{m-1}, \quad E(r - \beta)^2 = \frac{(\alpha^2 + \delta)}{(m-1)(m-2)}.$$

Consequently the bias of r is

$$\frac{\alpha}{m-1} - \frac{\alpha}{m} = \frac{\alpha}{m(m-1)}, \quad (9)$$

and the mean square error of r is

$$M(r) = E \left(r - \frac{\alpha}{m} - \beta \right)^2 \\ = E(r - \beta)^2 - \frac{2\alpha}{m} E(r - \beta) + \frac{\alpha^2}{m^2} \\ = \alpha^2 \left\{ \frac{1}{(m-1)(m-2)} - \frac{2}{m(m-1)} + \frac{1}{m^2} \right\} + \frac{\delta}{(m-1)(m-2)} \\ = \frac{\alpha^2(m+2)}{m^2(m-1)(m-2)} + \frac{\delta}{(m-1)(m-2)}. \quad (10)$$

As before, Quenouille's estimator is

$$t = \beta + \alpha \left\{ \frac{2}{x} - \frac{1}{2} \left(\frac{1}{x_1} + \frac{1}{x_2} \right) \right\} + \frac{u_1 + u_2}{x} - \frac{1}{2} \left(\frac{u_1}{x_1} + \frac{u_2}{x_2} \right),$$

so that

$$E(t - \beta) = \frac{\alpha(m-3)}{(m-1)(m-2)} - \frac{2\alpha}{m(m-1)(m-2)}. \quad (11)$$

Thus the bias is

$$E \left\{ \frac{2}{x} - \frac{1}{2} \left(\frac{1}{x_1} + \frac{1}{x_2} \right) \right\}^2 = E \left\{ \frac{4}{x^2} + \frac{1}{4} \left(\frac{1}{x_1^2} + \frac{1}{x_2^2} \right) - \frac{7}{2} \left(\frac{1}{x_1 x_2} \right) \right\}$$

As before,

$$E \left\{ \frac{u_1 + u_2}{x} - \frac{1}{2} \left(\frac{u_1}{x_1} + \frac{u_2}{x_2} \right) \right\}^2 = 2\delta E \left\{ \frac{2}{x^2} + \frac{1}{4} \left(\frac{1}{x_1^2} + \frac{1}{x_2^2} \right) - \frac{2}{x_1 x_2} \right\}.$$

and

Thus

$$E(t-\beta)^2 = \alpha^2 \left\{ \frac{4}{(m-1)(m-2)} + \frac{1}{2(m-2)(m-4)} - \frac{7}{2(m-2)^2} \right\} \\ + 2\delta \left\{ \frac{2}{(m-1)(m-2)} + \frac{1}{2(m-2)(m-4)} - \frac{2}{(m-2)^2} \right\} \\ = \frac{\alpha^2(m^2-8m+19)}{(m-1)(m-2)^2(m-4)} + \frac{\delta(m^2-7m+18)}{(m-1)(m-2)^2(m-4)}.$$

Consequently the mean square error of t is

$$M(t) = E(t-\beta)^2 - \frac{2\alpha}{m} E(t-\beta) + \frac{\alpha^2}{m^2} \\ = \frac{\alpha^2(m^2-5m^2+12m+16)}{m^2(m-1)(m-2)^2(m-4)} + \frac{\delta(m^2-7m+18)}{(m-1)(m-2)^2(m-4)}. \quad (12)$$

Comparing (10) and (12) we have

$$M(r) - M(t) = \frac{\alpha^2(m-16)}{m(m-1)(m-2)^2(m-4)} + \frac{\delta(m-10)}{(m-1)(m-2)^2(m-4)}. \quad (13)$$

Thus the mean square error of t is certainly less than that of r provided that $m > 16$ and might be less for other values of m between 10 and 16. Now in this formulation x has mean m and variance m . Thus the coefficient of variation of x is $m^{-\frac{1}{2}}$. We conclude that whenever the coefficient of variation of x is less than $\frac{1}{4}$, which will be satisfied by all except the most inaccurate estimators, Quenouille's estimator has a smaller mean square error than the ordinary ratio estimator. This is an exact result for any sample size. Coefficients of variation of greater than $\frac{1}{4}$ will generally result from smallness in sample size. For such cases an alternative method suggested by Quenouille (1956) will often be practicable and might well lead to a reduction in mean square error. The method can be described by considering the case of an estimator t based on n univariate observations. Let $t_1, \dots, t_n$ denote the estimators obtained by omitting the first, second, ..., n th observations in turn. Then $nt - (n-1) \sum_{i=1}^n t_i/n$ has bias of lower order than t and variance of the same order. Developments of this idea could be worked out for different types of ratio estimators but the matter will not be pursued here.

The analysis for ratios with Γ -type denominators has been carried out in terms of mean square error since the variance of t is found to be greater than that of r . Of course, an estimator which has both smaller bias and smaller mean square error than another is naturally preferable.

6. It may be of interest to quote the results obtained by applying the method to samples of two from the population of four pairs (x, y) used by Goodman & Hartley (1958) to compare a number of other types of estimators of the population mean of the y 's. The population values are (2, 2), (2, 6), (4, 6), (6, 10). Quenouille's estimator $\bar{y}_q$ and the four estimators considered by Goodman and Hartley, together with mean square errors, are:

Estimator	Mean square error
$\bar{y}_q = \frac{7}{2} \left(\frac{2(y_1+y_2)}{x_1+x_2} - \frac{1}{2} \left(\frac{y_1}{x_1} + \frac{y_2}{x_2} \right) \right)$	
$\tilde{y} = \frac{7}{2} \frac{y_1+y_2}{x_1+x_2}$	0.379
$\hat{y} = \frac{7}{4} \left(\frac{y_1}{x_1} + \frac{y_2}{x_2} \right)$	0.917
$y' = \frac{7}{4} \left(\frac{y_1}{x_1} + \frac{y_2}{x_2} \right) + \frac{3}{4} \left(y_1+y_2 - \frac{1}{2} \left(\frac{y_1}{x_1} + \frac{y_2}{x_2} \right) (x_1+x_2) \right)$	2.407
$\bar{y} = \frac{1}{2}(y_1+y_2)$	0.563
	2.667

In these formulae (x_1, y_1) and (x_2, y_2) are the sample values and $\frac{7}{2}$ is the population mean of the x 's. We see that $\bar{y}_q$ has a smaller mean square error than any of the estimators considered by Goodman & Hartley.

REFERENCES

- GOODMAN, L. A. & HARTLEY, H. O. (1958). *J. Amer. Statist. Ass.* **53**, 491-508.
 QUENOUILLE, M. H. (1949). *J. R. Statist. Soc. B*, **11**, 68-84.
 QUENOUILLE, M. H. (1956). *Biometrika*, **43**, 353-60.

On the probability integral transformation

By C. L. MALLOWS

University College London

1. Let $x_1, x_2, \dots, x_n$ be n random variables with a common probability density function $f(x, \theta_1, \theta_2, \dots, \theta_s)$ depending on s parameters, where

$$f(x, \theta_1, \theta_2, \dots, \theta_s) > 0 \quad (a < x < b). \quad (1)$$

If the parameters are known, the probability integral transformation defined by

$$z_i = \int_a^{x_i} f(t, \theta_1, \theta_2, \dots, \theta_s) dt \quad (a < x_i < b, i = 1, 2, \dots, n)$$

defines n random variables $z_1, z_2, \dots, z_n$ which are independent and are rectangularly distributed in $(0, 1)$. However, if we estimate the parameters by means of functions

$$F_1(x_1, \dots, x_n), \quad F_2(x_1, \dots, x_n), \quad \dots, \quad F_s(x_1, \dots, x_n)$$

of the observed values (which functions we may assume are not functionally dependent), and define n new random variables by

$$y_i = \int_a^{x_i} f(t, F_1, F_2, \dots, F_s) dt \\ = g(x_i, F_1, F_2, \dots, F_s), \quad \text{say} \quad (a < x_i < b, i = 1, 2, \dots, n),$$

then in general these variables will not be independent, nor will they be rectangularly distributed.

2. David & Johnson (1948) have remarked that the form of the joint distribution of the y 's will depend on the form of the distribution of the x 's, and may depend even on the values of the parameters in that distribution. Barton (1956) discussed these results while investigating Neyman's χ^2 -tests when the null hypothesis is composite, and Chernoff & Lehmann (1954) and Watson (1958) have considered related problems with reference to χ^2 -tests of goodness of fit. In these cases, the typical result is that the asymptotic distribution of the criterion considered is that of the sum of a χ^2 variable and one or more weighted squares of unit Normal variables; there is a partial loss of degrees of freedom relative to the case where the parameters are known. Barton shows by example that the amount of this loss may depend on the form of the null law, the particular estimators used, and even on the values of the parameters.

As a contribution to the study of the small-sample case, in the present note we show that the phenomenon observed by David & Johnson, that in some cases there are functional constraints between the y 's, is not at all general. On the basis of particular cases studied, David & Johnson conjectured that the rank of the Jacobian J (see (2)) will in general lie between $n-s$ and n . It will be shown that only in special cases will the rank be less than n . Notice that this loss of dimensionality does not correspond to the loss of degrees of freedom mentioned above.

3. Let

$$g_{i0} = g_0(x_i, F_1, \dots, F_s) = f(x_i, F_1, \dots, F_s), \\ g_{ir} = g_r(x_i, F_1, \dots, F_s) = \frac{\partial}{\partial F_r} g(x_i, F_1, \dots, F_s),$$

$$F_{ri} = \frac{\partial F_r}{\partial x_i}, \quad G_{ir} = \frac{g_{ir}}{g_{i0}}$$

$$r = 1, 2, \dots, s; \quad i = 1, 2, \dots, n.$$

for

(by (1), $g_{i0} > 0$ for $a < x_i < b$). Then we have

$$\frac{\partial y_i}{\partial x_i} = g_{i0} + \sum_{r=1}^s g_{ir} F_{ri} \quad (i = 1, 2, \dots, n), \\ \frac{\partial y_i}{\partial x_j} = \sum_{r=1}^s g_{ir} F_{rj} \quad (i, j = 1, 2, \dots, n; i \neq j).$$

The Jacobian matrix J of the transformation is given by

$$|J| = \left| \frac{\partial(y_1, y_2, \dots, y_n)}{\partial(x_1, x_2, \dots, x_n)} \right| = \prod_{i=1}^n g_{i0} |I_n + GF|, \quad (2)$$

where I_n is the unit $n \times n$ matrix, G is the $n \times s$ matrix (G_{ir}) and F is the $s \times n$ matrix (F_{ri}). Consider the partitioned $(n+s) \times (n+s)$ matrix

$$A = \begin{pmatrix} I_n + GF & 0 \\ 0 & I_s \end{pmatrix},$$

where I_s is the unit $s \times s$ matrix. We have

$$\rho(J) = \rho(I_n + GF) = \rho(A) - s.$$

But
$$\begin{pmatrix} I_n & 0 \\ F & I_s \end{pmatrix} \begin{pmatrix} I_n & G \\ 0 & I_s \end{pmatrix} \begin{pmatrix} I_n + GF & 0 \\ 0 & I_s \end{pmatrix} \begin{pmatrix} I_n & 0 \\ -F & I_s \end{pmatrix} \begin{pmatrix} I_n & -G \\ 0 & I_s \end{pmatrix} = \begin{pmatrix} I_n & 0 \\ 0 & I_s + FG \end{pmatrix} = B, \quad \text{say.}$$

All the multipliers on the left-hand side of this equation are non-singular; hence

$$n - \rho(J) = n + s - \rho(A) = n + s - \rho(B) = s - \rho(I_s + FG).$$

In general, $I_s + FG$ is non-singular, and so $\rho(J) = n$; only when some special relationship exists between f and the estimators $\{F_r\}$ will $\rho(J)$ be less than n . We remark that the form of this relationship does not seem to have anything to do with maximum likelihood. If $\rho(J) = n - s$, we have $\rho(I_s + FG) = 0$, i.e.

$$I_s + FG = 0$$

and conversely. Two lines of development for this case are now available. We may study either:

Problem (a), to find which forms for $f(x_i, \theta_1, \dots, \theta_s)$ satisfy these s^2 equations for given estimators $F_1, \dots, F_s$; or

Problem (b), to inquire what estimators $F_1, \dots, F_s$ should be used in any given case (i.e. given G_{ir}) to satisfy these equations.

4. *Problem (a)*. Suppose $s = 2$. For $\rho(J)$ to be $n - 2$ we require

$$1 + \sum_{i=1}^n \frac{g_{i1}}{g_{i0}} \frac{\partial F_1}{\partial x_i} = 0, \quad \sum_{i=1}^n \frac{g_{i2}}{g_{i0}} \frac{\partial F_1}{\partial x_i} = 0,$$

$$\sum_{i=1}^n \frac{g_{i1}}{g_{i0}} \frac{\partial F_2}{\partial x_i} = 0, \quad 1 + \sum_{i=1}^n \frac{g_{i2}}{g_{i0}} \frac{\partial F_2}{\partial x_i} = 0.$$

Now if, for example,

$$F_1 = \bar{x}, \quad F_2 = s = \frac{1}{\nu} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (\nu = n \text{ or } n-1),$$

$$F_{1i} = \frac{1}{n}, \quad F_{2i} = \frac{2}{\nu} (x_i - \bar{x})$$

we find the most general forms for G_{i1} and G_{i2} are

$$\frac{g_{i1}}{g_{i0}} = -1; \quad \frac{g_{i2}}{g_{i0}} = -\frac{x_i - \bar{x}}{2s}.$$

In order to find the most general form for g , we must now solve the simultaneous partial differential equations

$$\left. \begin{aligned} \frac{\partial g}{\partial x} + \frac{\partial g}{\partial \theta_1} &= 0, \\ \left(\frac{x - \theta_1}{2\theta_2} \right) \frac{\partial g}{\partial x} + \frac{\partial g}{\partial \theta_2} &= 0, \end{aligned} \right\}$$

the general solution of which (Piaggio, 1928, § 142) is

$$g(x, \theta_1, \theta_2) = h \left(\frac{x - \theta_1}{\sqrt{\theta_2}} \right) \quad (h \text{ arbitrary}).$$

For the case $s = 1$ we find in a similar fashion the solutions

$$F = \bar{x}:$$

$$g(x, \theta) = h(\phi(\theta)(x - \theta)) \quad (h, \phi \text{ arbitrary}),$$

$$F = \frac{1}{n} \sum_{i=1}^n \psi(x_i):$$

$$g(x, \theta) = h(\phi(\theta)(\psi(x) - \theta)) \quad (h, \phi \text{ arbitrary}),$$

$$F = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r:$$

$$g(x, \theta) = h \left(\frac{x}{\theta^{1/r}} + \phi(\theta) \right) \quad (h, \phi \text{ arbitrary}).$$

For the case $s = 3$ with

$$F_1 = \bar{x}, \quad F_2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \quad F_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3,$$

there is no possible form for g which will make $\rho(J) = n - 3$; it is, however, possible to reduce the rank by 2. We find the most general solution for which $\rho(J) = n - 2$ to be given by $g = h(\lambda)$ where

$$\lambda = \lambda(u, w) = \text{const.}$$

is an integral of

$$du + (\phi_1(w) + u\phi_2(w) + u^2\phi_3(w))dw = 0,$$

where $u = (x - \theta_1)/\theta_2^{1/2}$, $w = \theta_2^{1/2}/\theta_3^{1/2}$ and h , ϕ_1 , ϕ_2 and ϕ_3 are arbitrary functions.

5. Problem (b). If $s = 1$, and $g = h(x - \theta)$, we require

$$1 - \sum_{i=1}^n \frac{\partial F}{\partial x_i} = 0.$$

This will be satisfied if F is any (differentiable) function satisfying

$$F(x_1 + a, x_2 + a, \dots, x_n + a) = F(x_1, x_2, \dots, x_n) + a \quad (3)$$

for all allowable x and a . We note that this is the standard condition for the statistic F to be a 'measure of location'.

If $s = 2$ and $g = h\{(x - \theta_1)/\sqrt{\theta_2}\}$, we require

$$\begin{aligned} 1 - \sum_{i=1}^n \frac{\partial F_1}{\partial x_i} &= 0, & - \sum_{i=1}^n \frac{x_i - F_1}{2F_2} \frac{\partial F_1}{\partial x_i} &= 0, \\ - \sum_{i=1}^n \frac{\partial F_2}{\partial x_i} &= 0, & 1 - \sum_{i=1}^n \frac{x_i - F_1}{2F_2} \frac{\partial F_2}{\partial x_i} &= 0. \end{aligned}$$

These equations are satisfied if F_1 satisfies condition (3) and also $F_1(kx_1, kx_2, \dots, kx_n) = kF_1(x_1, x_2, \dots, x_n)$, while F_2 satisfies

$$\begin{aligned} F_2(x_1 + a, x_2 + a, \dots, x_n + a) &= F_2(x_1, x_2, \dots, x_n), \\ F_2(kx_1, kx_2, \dots, kx_n) &= k^2 F_2(x_1, x_2, \dots, x_n) \end{aligned}$$

for all allowable x , a and k . We note that these are the standard conditions for $\sqrt{F_2}$ to be a 'measure of dispersion'.

6. The above particular solutions cover a large proportion of the usual cases; there are, however, numerous important distributions which do not fall into any of the above forms; for example, the γ (or χ^2) distribution where the index parameter is estimated from the mean of the sample. In this case $\rho(J) = n$.

REFERENCES

- BARTON, D. E. (1956). Neyman's ψ_k^2 test of goodness of fit when the null hypothesis is composite. *Skand. Aktuar. Tidskr.* **39**, 216-45.
- CHERNOFF, HERMAN & LEHMANN, E. L. (1954). The use of maximum likelihood estimates in χ^2 tests for goodness of fit. *Ann. Math. Statist.* **25**, 579-86.
- DAVID, F. N. & JOHNSON, N. L. (1948). The probability integral transformation when parameters are estimated from the sample. *Biometrika*, **35**, 182-90.
- PIAGGIO, H. T. H. (1928). *Differential Equations*. London: G. Bell and Sons.
- WATSON, G. S. (1958). On Chi-square goodness-of-fit tests for continuous distributions. *J. R. Statist. Soc. B*, **20**, 44-61.

Extrema of quadratic forms with applications to statistics

By K. A. BUSH AND I. OLKIN*

University of Idaho and Michigan State University

Introduction and summary. In many statistical problems, one is frequently confronted with the determination of the extremum of a quadratic form or a ratio of quadratic forms, occasionally with some restrictions on the space. In this paper we consider a number of such problems: (i) stratified sampling, (ii) a moment problem, (iii) discrimination, (iv) canonical correlations, and show how the solutions may be obtained as a consequence of two inequalities. Standard results on Hotelling's T^2 , multiple correlation, and reliability may be similarly obtained.

* This work was supported in part by the Office of Ordnance Research, U.S. Army, while the author was on leave of absence at the University of Chicago.

Some examples. In what follows the notation $E(x)$ and $V(x)$ will denote the mean and variance, respectively.

(i) *Stratified sampling.* Let $X_{11}, \dots, X_{1N_1}, X_{21}, \dots, X_{2N_2}, \dots, X_{k1}, \dots, X_{kN_k}$, $N_1 + \dots + N_k = N$, be a finite population consisting of k strata, with mean $\bar{X} = \sum X_{ij}/N$ and intra-stratum variances S_i^2 , $i = 1, \dots, k$. A simple random sample $x_{i1}, \dots, x_{in_i}$ ($i = 1, \dots, k$) is chosen and an unbiased estimate of $\bar{X}$ is given by

$$\bar{x}_{st} = \sum N_i \bar{x}_i / N,$$

where $\bar{x}_i$ is the sample mean of the i th stratum. The variance of $\bar{x}_{st}$ is given by

$$V(\bar{x}_{st}) = \sum_{i=1}^k \frac{N_i^2 S_i^2}{N^2 n_i} - \sum_{i=1}^k \frac{N_i S_i^2}{N^2}.$$

The problem of optimum allocation is to determine $n_1, \dots, n_k$ so as to minimize $V(\bar{x}_{st})$ subject to the restriction $a_1 n_1 + \dots + a_k n_k = c$. If we let

$r = (r_1, \dots, r_k)$, $r_i = N_i S_i / N$, $D_a = \text{diag.}(a_1, \dots, a_k)$, $D_n = \text{diag.}(n_1, \dots, n_k)$, $e = (1, \dots, 1)$: $1 \times k$, then the problem is to find

$$\text{Min}_n r D_n^{-1} r' \quad \text{such that} \quad e D_n D_a e' = c.$$

(ii) *A moment problem.* Karush & Wolfsohn (1955) consider the problem of determining the minimum distance from the origin to a point $w = (w_1, \dots, w_k)$ in k -dimensional Euclidean space subject to the restrictions

$$\sum_1^k w_i = 1, \quad \sum_1^k i w_i = 1, \quad \dots, \quad \sum_1^k i^m w_i = 1,$$

for some $m < k-1$. If we let $e = (1, \dots, 1)$: $1 \times m+1$,

$$B = \begin{pmatrix} 1^0 & 2^0 & \dots & k^0 \\ 1^1 & 2^1 & \dots & k^1 \\ \dots & \dots & \dots & \dots \\ 1^m & 2^m & \dots & k^m \end{pmatrix},$$

then the problem becomes that of finding

$$\text{Min}_w w w' \quad \text{such that} \quad w B' = e.$$

(iii) *Discrimination.* Let $x = (x_1, \dots, x_p)$ follow a p -variate normal distribution with mean vector θ and covariance matrix Σ . Consider a linear combination of the variates, $d = w_1 x_1 + \dots + w_p x_p = w x'$; then $E(d) = w \theta'$ and $V(d) = w \Sigma w'$. The classical procedure is to choose the weight vector w so as to

$$\text{Max}_w \frac{(E d)^2}{V(d)} = \frac{(w \theta')^2}{w \Sigma w'}.$$

(iv) *Canonical correlation.* Let $y_1, \dots, y_q$, $x_1, \dots, x_p$ follow a $(p+q)$ -variate normal distribution with mean vector zero and covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix},$$

where $\Sigma_{11} = \text{cov}(y, y)$, $\Sigma_{12} = \text{cov}(y, x)$, $\Sigma_{22} = \text{cov}(x, x)$. Consider linear combinations

$$d_1 = u_1 y_1 + \dots + u_q y_q,$$

$$d_2 = v_1 x_1 + \dots + v_p x_p,$$

and choose the weight functions u and v to maximize the correlation coefficient between d_1 and d_2 . We note that $E d_1 d_2 = u \Sigma_{12} v'$, $V(d_1) = u \Sigma_{11} u'$, $V(d_2) = v \Sigma_{22} v'$, so that the problem is to find

$$\text{Max}_{u, v} \frac{u \Sigma_{12} v'}{\sqrt{(u \Sigma_{11} u')} \sqrt{(v \Sigma_{22} v')}}.$$

Two inequalities. A perusal of the problems indicate that they are of two types:

(I) Find $\text{Min}_w wAw'$ such that $wB = a$, where $w: 1 \times p$, $A: p \times p$ is positive definite, $B: p \times k$, $p \geq k$ is of rank k .

(II) Find $\text{Max}_w \frac{wAw'}{wBw'}$, where $A: p \times p$ is positive semi-definite, $B: p \times p$ is positive definite.

The usual solutions may involve Lagrange multipliers, direct differentiation, or some geometrical argument. The following inequalities permit solutions to (I) and (II).

Inequality 1. If M is positive definite, then

$$(xy')^2 \leq (xMx')(yM^{-1}y'), \quad (1)$$

with equality holding if and only if $xM = \alpha y$, where α is a scalar.

Proof. Since M is positive definite, there exists a factorization $M = LL'$, where L is non-singular. Using the Cauchy inequality $(uv')^2 \leq (uu')(vv')$ and the correspondence $u = xL$, $v' = L^{-1}y'$, we obtain (1). Equality holds if and only if $u = \alpha v$, i.e. if $xM = \alpha y$.

Alternatively we can now show that (1) is a special case of the generalized Cauchy inequality which states that in a linear space with inner product we have

$$|(f, g)| \leq \|f\| \cdot \|g\|,$$

where $\|f\| = (f, f)^{\frac{1}{2}}$ is the norm.

The space of all n -dimensional vectors is linear and the inner product may be chosen as xy' . Then if L is any $n \times n$ non-singular matrix and x and y vectors, we get

$$u = xL, \quad v = y(L^{-1}), \quad (u, v) = xy',$$

$$\|u\| = xLL'x' = xMx',$$

$$\|v\| = y(L^{-1})'L^{-1}y' = yM^{-1}y',$$

to obtain

$$(xy')^2 \leq (xMx')(yM^{-1}y').$$

By making the correspondence $w = x$, $A = M$, $aB' = y$, we obtain the bound in I,

$$(wAw') \geq \frac{(wBa')^2}{aB'A^{-1}Ba'} = \frac{(aa')^2}{aB'A^{-1}Ba'}, \quad (1')$$

with equality holding if and only if $wA = \alpha aB'$. Because of the constraint we also have

$$a = wB = \alpha aB'A^{-1}B, \quad \alpha = \frac{(aa')}{aB'A^{-1}Ba'},$$

so that equality holds if and only if

$$w = \frac{aa'}{aB'A^{-1}Ba'}(aB'A^{-1}).$$

Inequality 2. If A is positive semi-definite and B is positive definite, then

$$\theta_m \leq \frac{wAw'}{wBw'} \leq \theta_M, \quad (2)$$

where $\theta_1, \dots, \theta_p$ are the roots of $|A - \theta B| = 0$, $\theta_m = \min(\theta_1, \dots, \theta_p)$, $\theta_M = \max(\theta_1, \dots, \theta_p)$.

Proof. There exists a non-singular transformation $w = uQ$ such that

$$wAw' = uQAQ'u' = uD_\theta u',$$

$$wBw' = uQBQ'u' = uu',$$

where $D_\theta = \text{diag.}(\theta_1, \dots, \theta_p)$ and $\theta_1, \dots, \theta_p$ are the roots of $|A - \theta B| = 0$. Hence

$$\theta_m \leq \frac{wAw'}{wBw'} = \frac{uD_\theta u'}{uu'} \leq \theta_M.$$

We note that if A is of rank 1, then $A = a'a$, and there is one non-zero root of $|a'a - \theta B| = 0$, namely $aB^{-1}a'$, a fact which is apparent from (3) below. Thus in this case the two inequalities yield the same result

$$\frac{(wa')^2}{wBw'} \leq aB^{-1}a' = \theta_M.$$

At this point we can solve (i)–(iv). Problems (i), (ii) are solved using (1); (iii) may be solved using (1) or (2); (iv) is solved using (2).

(i) $(rD_n^{-1}r')(eD_a^{\frac{1}{2}}D_nD_a^{\frac{1}{2}}e') \geq (rD_a^{\frac{1}{2}}e')^2$, gives the lower bound, with equality holding if and only if $eD_a^{\frac{1}{2}}D_n = \alpha r$. Because of the restriction, the constant

$$\alpha = \frac{(eD_a^{\frac{1}{2}}D_nD_a^{\frac{1}{2}}e')}{rD_a^{\frac{1}{2}}e'} = \frac{c}{rD_a^{\frac{1}{2}}e'},$$

and hence

$$(n_1, \dots, n_k) = eD_n = \alpha rD_a^{-\frac{1}{2}} = \frac{crD_a^{-\frac{1}{2}}}{rD_a^{\frac{1}{2}}e'},$$

i.e.

$$n_j = \frac{cN_jS_j/\sqrt{a_j}}{\sum N_jS_j\sqrt{a_j}}.$$

(ii) The solution is given directly by (1') with $A = I$, $a = e: 1 \times m + 1$, and B replaced by B' :

$$ww' \geq \frac{(m+1)^2}{eBB'e'},$$

with equality holding if and only if

$$w = \frac{m+1}{eBB'e'}(eB).$$

(iii) Either (1) or (2) leads to the bound $\theta \Sigma^{-1} \theta'$.

(iv) Using (1) or (2) to obtain the first inequality and (2) to obtain the second inequality, we have

$$\frac{(u\Sigma_{12}v')^2}{(u\Sigma_{11}u')(v\Sigma_{22}v')} \leq \frac{u\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'u'}{u\Sigma_{11}u'} \leq \theta_M,$$

where θ_M is the maximum root of $|\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}' - \theta\Sigma_{11}| = 0$.

We note that the expression $aA^{-1}a'$ occurs as the solution of some of these problems, and we obtain an alternative expression

$$|M + x'y| = |M|(1 + yM^{-1}x'),$$

where M is non-singular and x, y and $1 \times n$ row vectors. This follows from the fact that $x'y$ is of rank 1, so that

$$|I + M^{-1}x'y| = 1 + \text{tr}(M^{-1}x'y) = 1 + yM^{-1}x'. \quad (3)$$

Thus if $x = y = a$, we have

$$aA^{-1}a' = \frac{|A + a'a|}{|A|} - 1.$$

REFERENCE

- KARUSH, W. & WOLFSOHN, N. Z. (1955). The distance to the origin of a certain point set in E^n . *Proc. Amer. Math. Soc.* 6, 323–32.

On certain properties of power-series distributions*

BY C. G. KHATRI
M. S. University of Baroda

INTRODUCTION

1. Noack (1950) has defined a power-series distribution as

$$P(\xi = x) = a_x Z^x / f(Z) \quad \text{for } x = 0, 1, 2, \dots \quad (1)$$

where $a_x Z^x \geq 0$, a_x is a function of x or constant, and $f(Z) = \sum_{x=0}^{\infty} a_x Z^x$ is convergent for $|Z| \leq r$. Here we establish the recurrence relations for cumulants and factorial cumulants, which are utilized to show that any power-series distribution in a single parameter is determined uniquely from its first two moments. The multivariate extensions of these results are shown with an illustration of multinomial distribution. All the above results extend to truncated power-series distributions.

* Read at the Second Reunion gathering (18 January 1959) of past and present students of the department of Statistics, University of Bombay, Bombay.

2. *Relation between factorial-cumulants and cumulants.* The r th factorial-moment is defined by $E\xi(\xi-h)(\xi-2h)\dots(\xi-rh+h)$, i.e. the coefficient of $\theta^r/r!$ in the expansion of $E(1+\theta h)^\xi/h$. Hence, the factorial-moment generating function is

$$F(\theta) = E(1+\theta h)^\xi/h. \quad (2)$$

Note that as $h \rightarrow 0$,

$$F(\theta) \rightarrow \phi(\theta) = E(e^{\theta\xi}), \quad (3)$$

the moment-generating function. The relations between the moment-generating function and factorial-moment generating function are

$$F(\theta) = \phi\left\{\frac{1}{h} \log(1+\theta h)\right\} \quad (4)$$

and

$$\phi(\theta) = F\{(e^{\theta h} - 1)/h\}. \quad (5)$$

Hence the relation between the factorial-cumulants and the cumulants can be established. If $\kappa_{[r]}$ is the r th factorial cumulant and κ_r is the r th cumulant, then

$$\kappa_{[r]} \rightarrow \kappa_r \quad \text{if } h \rightarrow 0. \quad (6)$$

3. *Properties of power-series distributions.*

(3.1). If $\kappa_{[j]}$ is the j th factorial cumulant of the power-series distribution (1), then

$$\kappa_{[j]} = Z^j h \left(Z^{-h+1} \frac{d}{dZ} \right)^j \log f(Z)$$

and

$$\kappa_{[j+1]} = Z \frac{d\kappa_{[j]}}{dZ} - jh \kappa_{[j]},$$

where

$$\left(Z^{-h+1} \frac{d}{dZ} \right)^j = \left(Z^{-h+1} \frac{d}{dZ} \right) \left(Z^{-h+1} \frac{d}{dZ} \right) \dots j \text{ times.}$$

Proof. It can be verified by induction that

$$(1+\theta h)^r Z^{-rh} \frac{\partial^r}{\partial \theta^r} \Phi\{Z(1+\theta h)^{1/h}\} = \left(Z^{-h+1} \frac{\partial}{\partial Z} \right)^r \Phi\{Z(1+\theta h)^{1/h}\}$$

and so, with $f(w)$ defined by (1),

$$\frac{\partial^r}{\partial \theta^r} \log f\{Z(1+\theta h)^{1/h}\} = (1+\theta h)^{-r} Z^{rh} \left(Z^{-h+1} \frac{\partial}{\partial Z} \right)^r \log f\{Z(1+\theta h)^{1/h}\}. \quad (7)$$

From (1), it is easy to see that the factorial-moment generating function $F(\theta)$ is $f\{Z(1+\theta h)^{1/h}\}/f(Z)$ and so

$$\begin{aligned} \kappa_{[r]} &= \left\{ \frac{\partial^r}{\partial \theta^r} \log F(\theta) \right\}_{\theta=0} \\ &= \left\{ \frac{\partial^r}{\partial \theta^r} \log f\{Z(1+\theta h)^{1/h}\} \right\}_{\theta=0}. \end{aligned} \quad (8)$$

Hence from (7), it follows immediately that

$$\kappa_{[r]} = Z^{rh} \left(Z^{-h+1} \frac{d}{dZ} \right)^r \log f(Z). \quad (9)$$

Differentiating (9) with respect to Z , we shall have the recurrence relation stated above.

COROLLARY 1. If $h = 0$, we have the relations for cumulants as

$$\kappa_j = \left(Z \frac{d}{dZ} \right)^j \log f(Z) = Z \frac{d\kappa_{j-1}}{dZ}. \quad (10)$$

COROLLARY 2. If $h = 1$, we have the relations as

$$\kappa_{[j]} = Z^r \frac{d^r}{dZ^r} \log f(Z) = Z \frac{d\kappa_{[j-1]}}{dZ} - (j-1)\kappa_{[j-1]}. \quad (11)$$

(3.2). The power-series distribution is uniquely determined from its first two cumulants (or moments).

Proof. Let the power-series which determines a distribution be given by $f(Z)$ and the first two cumulants be given in the parameter p where $Z = g(p)$ is unknown. Then from the cumulant relation (10),

$$\text{we have} \quad \kappa_1 = Z \frac{d}{dZ} \log f(Z) = \frac{d}{dp} \log y(p) \bigg/ \frac{d \log Z}{dp}, \quad \text{where } y(p) = f(g(p)),$$

and

$$\kappa_2 = Z \frac{d\kappa_1}{dZ} = \frac{d\kappa_1}{dp} \bigg/ \frac{d \log Z}{dp}.$$

Therefore

$$\frac{d \log Z}{dp} = \frac{d\kappa_1}{dp} / \kappa_2 = \frac{d}{dp} \log y(p) / \kappa_1. \quad (12)$$

Hence

$$\log Z = \int \left(\frac{d\kappa_1}{dp} / \kappa_2 \right) dp + c_3 = \log g_1(p) + \log c_1, \quad \text{say,} \quad (13)$$

i.e.

$$Z = c_1 g_1(p)$$

and

$$\log y(p) = \int \left(\kappa_1 \frac{d\kappa_1}{dp} / \kappa_2 \right) dp + c_4 = \log y_1(p) + \log c_2, \quad \text{say,} \quad (14)$$

i.e.

$$y(p) = c_2 y_1(p), \quad (14)$$

where c_1, c_2, c_3 and c_4 are constants.

Let a_x be the coefficient of $\{g_1(p)\}^x$ in the expansion of $y_1(p)$, then $c_2 c_1^{-x} a_x$ will be the coefficient of Z^x in the expansion of $y(p)$. Hence, the probability that the random variable ξ takes the value x is

$$c_2 c_1^{-x} a_x Z^x / y(p) = a_x \{g_1(p)\}^x / y_1(p). \quad (15)$$

Hence, without loss of generality, we can assume in equations (13) and (14) that $c_1 = c_2 = 1$, i.e.

$$Z = \exp \left\{ \int \left(\frac{d\kappa_1}{dy} / \kappa_2 \right) dy \right\}$$

and

$$f(Z) = \exp \left\{ \int \left(\kappa_1 \frac{d\kappa_1}{dy} / \kappa_2 \right) dy \right\}. \quad (16)$$

Now suppose that any two distributions have κ_1 and κ_2 equal, then by the expressions derived in (15) and (16), the two distributions are equal. Hence we have proved the theorem.

Note that two successive cumulants of a power-series distribution other than the first two cumulants cannot determine the distribution uniquely.

4. Illustrative examples.

(a) Negative binomial distribution.

Let $f(Z) = (1-Z)^{-n}$ for $0 \leq Z \leq 1, n > 0, Z = p/(1+p) = p/q, p \geq 0$. Then

$$P(\xi = x) = \frac{\Gamma(n+x)}{x! \Gamma(n)} p^x q^{-n-x} \quad \text{for } x = 0, 1, 2, \dots \quad (17)$$

We have from cumulant relation

$$\kappa_r = Z \frac{d\kappa_{r-1}}{dZ} = pq \frac{d\kappa_{r-1}}{dp},$$

and so $\kappa_1 = np, \kappa_2 = npq, \kappa_3 = npq(q+p)$ and $\kappa_4 = npq(1+6pq)$. Conversely, if κ_1 and κ_2 are given, we have from (16)

$$Z = \exp \left(\int \frac{1}{pq} dp \right) = \frac{p}{q}$$

and

$$f(Z) = \exp \left(\int \frac{n}{q} dp \right) = q^n = (1-Z)^{-n}.$$

Hence we have the negative binomial distribution.

(b) (i) Truncated negative binomial distribution truncated at $x = 0$.

Here $f(Z) = (1-Z)^{-n} - 1$ for $0 \leq Z \leq 1, n \geq 0, Z = p/q, p \geq 0$. Then

$$P(\xi = x) = \frac{\Gamma(n+x)}{x! \Gamma(n)} \frac{p^x q^{-x}}{q^n - 1} \quad \text{for } x = 1, 2, \dots \quad (18)$$

In this case also, we have $\kappa_r = pq(d\kappa_{r-1}/dp)$, and so

$$\kappa_1 = npq^n/(q^n - 1), \quad \kappa_2 = npq^n(q^{n+1} - q - np)/(q^n - 1)^2, \quad \text{etc.}$$

Conversely, from κ_1 and κ_2 , we can obtain $Z = p/q$ and $f(Z) = q^n - 1$, i.e. $f(Z) = (1-Z)^{-n} - 1$.

(ii) Particular case of (i). In (i), let $n \rightarrow 0$. Then we shall have

$$\kappa_1 = p/\log q \quad \text{and} \quad \kappa_2 = p(q \log q - p)/(\log q)^2. \quad (19)$$

From these cumulants, we have

$$Z = \exp \left(\int \frac{dp}{pq} \right) = \frac{p}{q} \quad \text{and} \quad f(Z) = \exp \left(\int \frac{dp}{q \log q} \right) = \log q,$$

i.e.

$$f(Z) = -\log(1-Z).$$

Hence the distribution is the logarithmic series

$$P(\xi = x) = p^x q^{-x} / x \log q \quad \text{for } x = 1, 2, \dots \quad (20)$$

(c) *Positive binomial distribution.*

Let $f(Z) = (1+Z)^n$, $Z \geq 0$, n is a positive integer, $Z = p/(1-p)$, $0 \leq p \leq 1$. Then

$$P(\xi = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x = 0, 1, \dots, n. \quad (21)$$

The cumulant relation is

$$\kappa_r = Z \frac{d\kappa_{r-1}}{dZ} = p(1-p) \frac{d\kappa_{r-1}}{dp},$$

and so

$$\kappa_1 = np \quad \text{and} \quad \kappa_2 = np(1-p).$$

The converse problem is similar to (a) and (b).

(d) *Poisson distribution.* This can be treated in the same manner.

5. *Multivariate extension with multinomial distribution as an illustration.*

Let

$$f(Z_1, Z_2, \dots, Z_k) = \sum_{x_1, \dots, x_k} a_{x_1, \dots, x_k} Z_1^{x_1} Z_2^{x_2} \dots Z_k^{x_k}$$

be a convergent series such that $a_{x_1, \dots, x_k}$ may be a function of $x_1, \dots, x_k$ or constant and

$$a_{x_1, \dots, x_k} Z_1^{x_1} \dots Z_k^{x_k} \geq 0.$$

Then we shall define the probability of the random variables $\xi_1, \dots, \xi_k$ taking for values $x_1, \dots, x_k$, respectively, as

$$P(\xi_1 = x_1, \dots, \xi_k = x_k) = a_{x_1, \dots, x_k} Z_1^{x_1} \dots Z_k^{x_k} / f(Z_1 \dots Z_k). \quad (22)$$

(A) *Example.* Let $f(Z_1, \dots, Z_k) = (1 + Z_1 + \dots + Z_k)^n$, where n is a positive integer,

$$Z_i = p_i / p_{k+1}, \quad 0 \leq p_i \leq 1, \quad \sum_{i=1}^{k+1} p_i = 1.$$

Then

$$P(\xi_1 = x_1, \dots, \xi_k = x_k) = \frac{n!}{x_1! \dots x_k! x_{k+1}!} p_1^{x_1} \dots p_k^{x_k} p_{k+1}^{x_{k+1}},$$

where

$$\sum_{i=1}^{k+1} x_i = n \quad (x_i = 0, 1, 2, \dots, n). \quad (23)$$

This is a positive multinomial distribution.

Let $\kappa_{[r_1, \dots, r_k]}$ be the factorial cumulant of the distribution given in (22). Then it is easy to show that

$$\kappa_{[r_1, \dots, r_k]} = \prod_{i=1}^k Z_i^{r_i} h_i \left(Z_i^{-h_i+1} \frac{\partial}{\partial Z_i} \right)^{r_i} \log f(Z_1, \dots, Z_k), \quad (24)$$

or

$$\kappa_{[r_1, \dots, r_{i-1}, r_i+1, r_{i+1}, \dots, r_k]} = Z_i \frac{\partial \kappa_{[r_1, \dots, r_k]}}{\partial Z_i} - r_i h_i \kappa_{[r_1, \dots, r_k]}.$$

If $h_i = 0$ for all i in (24), we can at once reach the corresponding relations for cumulants.

Example. Let

$$f(Z_1, Z_2, \dots, Z_k) = (1 + Z_1 + \dots + Z_k)^n, \quad \text{where } Z_i = p_i / p_{k+1}, \quad 0 \leq p_i \leq 1, \quad \sum_{i=1}^{k+1} p_i = 1.$$

We shall here only write down the means, variances and covariances; other cumulants can be written down from the recurrence relations.

$$E(\xi_i) = n Z_i / \left(1 + \sum_{j=1}^k Z_j \right) = n p_i \quad \text{for } i = 1, 2, \dots, k,$$

$$V(\xi_i) = Z_i \frac{\partial}{\partial Z_i} E(\xi_i) = n p_i (1 - p_i) \quad \text{for } i = 1, 2, \dots, k,$$

and

$$\text{cov}(\xi_i, \xi_j) = Z_i \frac{\partial}{\partial Z_i} E(\xi_j) = Z_j \frac{\partial}{\partial Z_j} E(\xi_i) = -n p_i p_j \quad \text{for } i \neq j = 1, 2, \dots, k. \quad (25)$$

Now consider the converse problem of obtaining the distribution from means, variances and covariances given in parameters $(y_1, y_2, \dots, y_k)$ if the distribution is of the power-series type.

Let Z_i be the function of $(y_1, y_2, \dots, y_k)$, $i = 1, 2, \dots, k$. Then

$$\begin{aligned} \text{cov}(\xi_i, \xi_j) &= Z_j \frac{\partial}{\partial Z_j} E(\xi_i) = Z_j \sum_{t=1}^k \frac{\partial y_t}{\partial Z_j} \frac{\partial}{\partial y_t} E(\xi_i) \\ &= Z_i \frac{\partial}{\partial Z_i} E(\xi_j) = Z_i \sum_{t=1}^k \frac{\partial y_t}{\partial Z_i} \frac{\partial}{\partial y_t} E(\xi_j) \end{aligned}$$

for $i \neq j = 1, 2, \dots, k$ and

$$V(\xi_i) = Z_i \sum_{t=1}^k \frac{\partial y_t}{\partial Z_i} \frac{\partial}{\partial y_t} E(\xi_i) \quad (i = 1, 2, \dots, k). \quad (26)$$

From the set of equations (26), we can obtain the solution for $Z_i(\partial y_i / \partial Z_i)$, as a function $g_{i,t}(y_1, \dots, y_k)$ ($i, t = 1, 2, \dots, k$). Then we have

$$dy_i = g_{i,1}(y_1 \dots y_k) \frac{dZ_1}{Z_1} + \dots + g_{i,k}(y_1, \dots, y_k) \frac{dZ_k}{Z_k} \quad \text{for } i = 1, 2, \dots, k. \quad (27)$$

From the above set of equations, we can obtain dZ_i/Z_i or $(1/Z_i)(\partial Z_i / \partial y_i)$ in terms of functions of $(y_1, y_2, \dots, y_k)$ and hence obtain Z_i . After this operation, we can express $E(\xi_i)$ in terms of $Z_1, Z_2, \dots, Z_k$ and obtain the solution from $Z_i(\partial / \partial Z_i) \log f(Z_1, \dots, Z_k) = E(\xi_i)$. Thus we can determine the power series, and so the distribution.

For example, let the means, variances and covariances be given as (25). Then

$$\text{cov}(\xi_i, \xi_j) = -np_i p_j = nZ_j \frac{\partial p_i}{\partial Z_j} = nZ_i \frac{\partial p_j}{\partial Z_i} \quad \text{by the use of (26),}$$

i.e.

$$Z_j \frac{\partial p_i}{\partial Z_j} = Z_i \frac{\partial p_j}{\partial Z_i} = -p_i p_j \quad \text{for } i \neq j = 1, 2, \dots, k.$$

Similarly

$$Z_i \frac{\partial p_i}{\partial Z_i} = p_i(1 - p_i) \quad \text{for } i = 1, 2, \dots, k.$$

Hence

$$\frac{dp_i}{p_i} = -p_1 \frac{dZ_1}{Z_1} - p_2 \frac{dZ_2}{Z_2} - \dots - p_{i-1} \frac{dZ_{i-1}}{Z_{i-1}} + (1 - p_i) \frac{dZ_i}{Z_i} - p_{i+1} \frac{dZ_{i+1}}{Z_{i+1}} - \dots - p_k \frac{dZ_k}{Z_k}. \quad (28)$$

To obtain dZ_i/Z_i in terms of dp_i/p_i ($i = 1, 2, \dots, k$), we have to obtain the inverse of

$$\begin{pmatrix} 1-p_1 & -p_2 & \dots & -p_k \\ -p_1 & 1-p_2 & \dots & -p_k \\ \dots & \dots & \dots & \dots \\ -p_1 & -p_2 & \dots & 1-p_k \end{pmatrix} = I - \mathbf{1p}',$$

where $\mathbf{p}' = (p_1, p_2, \dots, p_k)$, $\mathbf{1}' = (1, 1, \dots, 1)$ and I is a $k \times k$ identity matrix. We can easily show that

$$(I - \mathbf{1p}')^{-1} = I + \mathbf{1p}'/p_{k+1}, \quad p_{k+1} = 1 - \sum_{j=1}^k p_j.$$

Thus the solution is

$$\frac{1}{Z_i} \frac{\partial Z_i}{\partial p_i} = \left(1 + \frac{p_i}{p_{k+1}}\right) / p_i \quad \text{and} \quad \frac{1}{Z_i} \frac{\partial Z_i}{\partial p_t} = 1/p_{k+1} \quad \text{for } i \neq t = 1, 2, \dots, k.$$

Hence we have $\log Z_i = \log p_i/p_{k+1}$, i.e. $Z_i = p_i/p_{k+1}$. From

$$E(\xi_i) = nZ_i \left/ \left(1 + \sum_{j=1}^k Z_j\right) \right. = Z_i \frac{\partial}{\partial Z_i} \log f(Z_1, \dots, Z_k),$$

we obtain

$$\log f(Z_1, \dots, Z_k) = n \log \left(1 + \sum_{j=1}^k Z_j\right) + c(Z_1, \dots, Z_{i-1}, Z_{i+1}, \dots, Z_k), \quad (29)$$

where $c(Z_1, \dots, Z_{i-1}, Z_{i+1}, \dots, Z_k)$ is the constant of integration. Since (29) is true for all i , we must have

$$\log f(Z_1, \dots, Z_k) = n \log \left(1 + \sum_{j=1}^k Z_j\right), \quad \text{i.e. } f(Z_1, \dots, Z_k) = \left(1 + \sum_{j=1}^k Z_j\right)^n.$$

Hence, we shall have the probability of the random variables $\xi_1, \xi_2, \dots, \xi_k$ to take the values $x_1, x_2, \dots, x_k$ as given in (23).

(B) *Negative multinomial distribution.*

This is generated by $f(Z_1, Z_2, \dots, Z_k) = (1 - Z_1 - Z_2 - \dots - Z_k)^{-n}$, where

$$Z_i = p_i \left/ \left(1 + \sum_{j=1}^k p_j\right) \right., \quad 0 \leq Z_i \leq 1, \quad \sum_{j=1}^k Z_j \leq 1 \quad \text{and} \quad p_i \geq 0.$$

Then

$$P(\xi_1 = x_1, \dots, \xi_k = x_k) = \frac{\Gamma\left(n + \sum_{j=1}^k x_j\right)}{x_1! \dots x_k! \Gamma(n)} p_1^{x_1} \dots p_k^{x_k} \left(1 + \sum_{j=1}^k p_j\right)^{-\left(n + \sum_{j=1}^k x_j\right)} \quad \text{for all } x_i = 0, 1, 2, \dots \quad (30)$$

We shall have means, variances and covariances as

$$E(\xi_i) = np_i, \quad V(\xi_i) = np_i(1 + p_i) \quad \text{and} \quad \text{cov}(\xi_i, \xi_j) = np_i p_j \quad \text{for } i \neq j = 1, 2, \dots, k.$$

The converse result can also be obtained in the same manner as the above positive multinomial one.

REFERENCE

NOACK, A. (1950). A class of random variables with discrete distributions. *Ann. Math. Statist.* **21**, 127-32.

REVIEWS

The Advanced Theory of Statistics. Volume I, Distribution Theory. By MAURICE G. KENDALL and ALAN STUART. London: Charles Griffin and Co. Ltd. 1958. Pp. xii + 433. 84s.

The 16 years which have elapsed since the publication in 1943 of vol. I of M. G. Kendall's *The Advanced Theory of Statistics* have witnessed a remarkable, indeed a rather bewildering output of text-books on statistics of very varied quality. It is therefore most satisfactory that Professor Kendall, with Mr Alan Stuart's help, has already completed part of the process of bringing his pioneer contribution up to date. To maintain the same comprehensive standard, three volumes will be required in place of two; the first was published in the summer of 1958 and the second is promised in 1960.

Broadly speaking, the first 13 chapters of the new vol. I correspond to the first 11 chapters of the original vol. I. The increase in number results from dividing the old chapter 7 on Probability and Likelihood into chapters on The Calculus of Probabilities and Probability and Statistical Inference, and the splitting of chapter 11 on Approximations to Sampling Distributions into two.

The expansion of the work into three volumes has no doubt made it possible for the authors to think out again the order and form of presentation of the later material; as a result the present volume closes with three in place of the original five chapters; of these, chapter 14 on Order-statistics and 15 on The Multivariate Normal Distribution and Quadratic Forms are new, while the final chapter 16 on Distributions Associated with the Normal contains material drawn together from elsewhere in the original vols. I and II.

The changes made in turning the old chapters into the new are of two kinds. In the first place there are additions of entirely new material made necessary by advances in statistical theory and improvements in its techniques. Thus we find, for example, accounts of R. A. Fisher's logarithmic distribution with an example (§§ 5.16, 5.17); of the log-normal distribution and of N. L. Johnson's S_U and S_B frequency curves (§§ 6.27-6.35); a reference to uses of the method of steepest descents (§ 11.13), and a brief account of the method of deriving sampling moments systematically with the help of tables of symmetric functions, such as those recently compiled by F. N. David and M. G. Kendall (§ 12.5). But apart from the introduction of new material, it is evident that the authors have reconsidered the detailed presentation throughout and where they thought that this would clarify the argument have made alterations by slight re-arrangement of the text, by addition of paragraphs and by improvement of the illustrative examples. This process of improvement has been helped by changes in typography — the main section headings are now in bold instead of italic type, the tables are headed in bold type and printed in a more pleasing manner and the algebraic formulae are now displayed more compactly.

The first doubling of chapters mentioned above has permitted an improved presentation of the brief introduction to the ideas of probability, likelihood and inference which was inserted in the original plan at this half-way stage to give point to the development in vol. I of what is primarily a mathematical theory of statistical distributions. Thus the concept of a random variable is now treated in greater detail (§§ 7.12-7.19). It is interesting to note the authors' conclusion, summed up in the following sentences (p. 180), that there is no unique best method to relate a mathematical theory to the way we think:

'It seems, however, that every man must choose for himself and that his psychological make-up, his experience and his fields of interest all determine the kind of axiomatization which he prefers. In statistics it is a mark of immaturity to argue overmuch about the fundamentals of probability theory.'

In the last sentence the intended emphasis is presumably on 'overmuch' and the authors are thinking of that kind of dogmatic assertion which claims that one method of approach is right and the others wrong. They can hardly mean that there should not be continued interest in the fundamentals nor fail to realise that when two or three people discuss such matters keenly there is bound to be argument!

The second doubling of chapters, 11 into 12 and 13, makes possible some re-arrangement and expansion in the handling of sampling moments and sampling cumulants, including the sampling from finite populations. As now presented, the second of the two chapters deals with bivariate k -statistics and the proofs of the combinatorial rules stated in the earlier chapter.

The chapter on Order Statistics (14), besides incorporating the sections on the distribution of m th values and of range which previously appeared in the chapter on Standard Errors, contains a good

deal of new matter. The new chapter 15 on The Multivariate Normal Distribution and Quadratic Forms introduces the vector and matrix notation which presumably will be taken up again in vol. III where the application of methods of multivariate analysis is to be included.

The original vol. I closed with a mixed series of chapters in which, besides the development of the necessary theory, some straightforward numerical illustrations were given of the uses (a) of the χ^2 test, (b) of bivariate correlation and regression, and (c) of other measures of association. All this illustration of application has now been removed, presumably to be included in the later volumes, and instead we have here brought together in a single chapter on Distributions Associated with the Normal (16), the mathematical properties of and approximations to the distributions of the standard normal-theory statistics χ^2 , t , F and z , r and b . There are obvious advantages in this arrangement which makes available in one place the distribution theory of the tools required for later application.

The volume closes with 12 single-page tables, 10 dealing with the normal distribution, χ^2 , t , F and z and two giving (i) augmented symmetric functions in terms of power-sums up to weight 6 and (ii) multiple k -statistics in terms of augmented symmetric functions up to 6th order. The last two tables are linked with the discussion in chapter 12. There is also a list of References to the authors quoted and a 14 pp. Index. It is a pity that here (as well as in another recent book on statistics) the name Gosset has been spelt with two t 's!

The Exercises left for solution by the reader have been greatly increased in number so that there are now about 20 at the end of each chapter. This should prove a valuable addition for the serious student.

It is of course impossible at present for a reviewer to survey this great undertaking as a whole, because its successful rounding off must depend on the way in which the later volumes on *Statistical Inference and Statistical Relationship* (2) and *Statistical Planning and Analysis, and Time-Series* (3) are built up on the present volume on *Distribution Theory*. With a project of this magnitude there are inevitably alternative methods of arranging the order of the parts. In dealing with statistics and probability the writer of an expository text is faced with the problem of ordering in a single dimension, subjects whose links are essentially multi-dimensional. From the teaching point of view it is instructive in itself to see how the present authors have attacked their problem.

The decision on the broad plan was taken by Professor Kendall nearly 20 years ago, but the changes already appearing in the revised vol. I suggest that this plan is being modified to some extent. The removal of much of the numerical illustration from the later chapters, while retaining a generous use of such illustration in the first few chapters appears a sound policy. A more difficult matter for decision must have been where, how and to what extent to introduce the ideas of inference and random sampling. The placing of an account of Bayes's Theorem and Maximum Likelihood with a discussion on randomness and bias in practical sampling in the middle of vol. I is undoubtedly in some ways anomalous. So, too, is the inclusion in this volume of tables of percentage points of χ^2 , t , F and z before the reader has been shown how to use them. But if it is remembered that ultimately all three volumes will be available together so that the reader can choose at will his own linking and cross-linking of the parts, criticism on such points is hardly valid.

It remains only for the reviewer to remark that he awaits the appearance of vols. II and III in pleasant anticipation and that at the same time he wishes there were a solution to the problem of how the advanced student is to find means of purchasing his own copies of the complete series at a total price which is likely to approach £12.

E. S. PEARSON

Planning of Experiments. By D. R. Cox. New York: John Wiley and Sons Inc.; London: Chapman and Hall Ltd. 1958. Pp. vii + 308. 60s.

Let us define our terms. A statistician (i.e. a real statistician) is a person experienced in handling data, in all the various stages from planning to presentation (decision-making is a further step which usually involves non-statistical considerations). A mathematical statistician has some fluency in various techniques involved in these stages, and is usually actively engaged in extending these techniques, or in inventing new ones. A statistical mathematician studies various idealized problems arising out of the experience of real statisticians—he is not immediately interested in real data. All three classifications may apply to one individual at different times. The statistician of most immediate value to a commercial employer is the real mathematical variety.

It is a common and valid criticism of many university departments and standard texts that they omit sufficient mention of the many difficulties sure to be encountered in applying the mathematical techniques in real life, and offer little guidance in these matters. Thus new graduates very seldom qualify

to be called real statisticians, and it usually takes them some time to supplement their formal training with practical experience.

Dr Cox's book is extremely welcome in that it concentrates on just those matters where little guidance is available elsewhere. The practical aspects of experimental design are here dealt with in a highly individual way—the discussion avoids mathematical technicalities, yet is careful and detailed; it is mostly in terms of concrete examples, taken from many different fields. It is intended primarily for the private reading and reference of the experimental workers, and requires no specialized knowledge for its understanding; thus for instance the details of the analyses of variance involved receive only incidental mention.

Of course, there is no substitute for experience; and Dr Cox is careful to point out that his examples have often been simplified so that the principle under discussion might more clearly be brought out; references to the original accounts have been given wherever possible. Nevertheless, the book goes much farther than is usual in either elementary or advanced texts in bringing out the difficulties, and in expounding general principles. Its relation to such works as that of Cochran and Cox is not so much introductory as complementary.

The first nine chapters deal with basic concepts and the key designs; the last five touch briefly on some more advanced topics. The concern is solely with comparative experiments so that, for example, it does not cover the planning of surveys.

In summary, this is a book about real statistics. It is easy to predict that it will have considerable success; it should be made required reading for all students of statistics.

C. L. MALLOWS

Theory and Methods of Scaling. By WARREN S. TORGERSON. New York: John Wiley and Sons Inc.; London: Chapman and Hall, Ltd. 1958. Pp. xiv + 460. 76s.

In 1950 the American Social Sciences Research Council appointed a committee to study the relative merits of different methods of scaling; and Dr Torgerson was chosen to review and summarize, under the general guidance of the committee, the material available. He begins with a theoretical examination of the different types of scientific measurement, and then proceeds to classify the various techniques for metrical scaling which have been elaborated during recent years. After discussing the alternative criteria that could be used for such a classification, he develops a working scheme based on 'the different ways in which the variations of the subjects' responses may be allocated'. This leads him to distinguish three main groups.

The first group of procedures consists of those which attribute the variations in the subjects' responses to individual differences in the subjects themselves: this he terms the 'subject-centred approach'. It is the principle underlying most of the work on mental testing. But, largely because it has been so fully discussed elsewhere, Dr Torgerson has decided to exclude it from his own review. The second group adopts what he calls the 'stimulus-centred approach'. Here the systematic variations in the subjects' reactions are attributed to the way the stimuli vary in respect of some specific attribute. The majority of the well-known psychophysical methods fall into this group; and these with their latest modifications are discussed in considerable detail. Separate chapters are devoted to methods based on the 'law of comparative judgement', the 'law of categorical judgement', and to the theory of multidimensional scaling—a topic to which Dr Torgerson has himself made important contributions.

The third group adopts the 'response method'. Here the variability in the subjects' reactions is ascribed partly to variations in the subjects and partly to variations in the stimuli. This, as Dr Torgerson points out, is the most general method of all; but, owing to the complexity of the problems involved, the methods devised are relatively few and comparatively recent. He recognizes two main subclasses—'deterministic models', of which Guttman's is the best known example, and 'probabilistic models', which include Lazarsfeld's method of 'latent structure' and Coombs's 'unfolding technique'.

All the more important procedures falling under each of these several heads are lucidly explained with detailed illustrations; the underlying theories on which they rest are critically examined; and their use and limitations impartially discussed. Much of the material included has hitherto been available only in isolated articles or specialist text-books. Hence the volume will be invaluable to every investigator who seeks to apply quantitative techniques to social or psychological problems.

CYRIL BURT

Linear Programming: Methods and Applications. By SAUL I. GASS. New York and London: McGraw-Hill Book Co. Inc. 1958. Pp. 223. 50s. 6d.

This book differs from most other books on Linear Programming in that it contains an extensive formal exposition of the underlying theory. It answers in this respect a real need (even if it is not the first to attempt it, as the publishers' blurb claims). The text had been prepared for an academic course, and is thus addressed to an audience—and readership—with well-defined and predictable mathematical knowledge. For British readers the level of sophistication is perhaps best defined by mentioning that matrix notation is used throughout, but that about 20 pages are devoted to the formal theory of matrices, convex sets and simultaneous linear equations.

After this, there follows an equally readable part II—Methods: Theoretical and Computational. This contains seven chapters, all of subjects related to the Simplex Method; it includes a chapter on Parametric Linear Programming (due partly to the author himself), which deals with a minor but important aspect of programming under uncertainty. There is, also, a survey of available automatic programmes for digital computers, necessarily and admittedly already incomplete.

However, this is not the only feature which makes the book appear somewhat dated. Stochastic, non-linear, and discrete Linear Programming are mentioned only casually, clearly because some newer methods were unknown when the book was conceived. But even regarding the foundations there exist now—at any rate in the reviewer's opinion—methods of higher didactic value than those presented by the author. He has followed here closely the original publications, whose importance is, in many cases, merely historical. Even so, the purpose of making the matter clearly understood is achieved, and this is praise which some other books on the same subject have failed to deserve.

When we come to part III—Applications, we find the author less successful. He gives here a rather scrappy list of descriptions of Linear Programming in practice, but it is impossible for a reader who does not know the examples already to guess how some of them could be put into Linear Programming form. (This applies, with particular force, to the Paper Trim Problem, on p. 182, and to the Travelling Salesman Problem, on p. 186.) It is true that the book contains a bibliography of applications, grouped by subjects such as agricultural, industrial, military, etc., but too many of the titles refer to reports of U.S. Government agencies, even where their text has since appeared in readily available periodicals, e.g. Management Science or Operations Research.

The last chapter contains a clearly and well-written description of two-person zero-sum games and establishes their connexion with Linear Programming.

The proof reading has been done very carefully: only very few misprints survived (e.g. coeds for codes on p. 130; also, the enumeration on p. 6 should be one of alternatives, not of simultaneous features).

In spite of the objections raised above, the book is a very useful contribution to the explanatory literature, and an essential item on the shelf of any Operational Research worker who is interested in the foundations as well as the applications of his subject.

S. VAJDA

Wahrscheinlichkeitsrechnung und Mathematische Statistik (Probability and Mathematical Statistics). Second edition. By M. FISZ. Berlin: VEB Deutscher Verlag der Wissenschaften (Hochschulbücher für Mathematik Band 40). 1958. Pp. 528. DM. 14.

The first edition of this book (*Rachunek prawdopodobieństwa i statystyka matematyczna*) was reviewed in *Mathematical Reviews*, 1955, 16, 492 by Z. W. Birnbaum. This second edition, much more extensive and complete, appeared in 1958, simultaneously in Polish and German.

The book consists of two parts: the first is concerned with the theory of probability, the second with statistical theory, in particular the theory relating to inference from sample values to the universe from which the sample has been drawn. In the first five chapters the author discusses basic ideas: random events, the axiomatic theory of probability, random variates and their distributions and characteristic functions in one- and multi-dimensional spaces. The sixth chapter is devoted to the more advanced applied fields, and clear intuitive interpretations of many of the mathematical formulae help to make this book easier to read. It should be noted that the author uses the expression 'the most efficient estimator' for the estimator with minimum variance instead of the more usual definition 'the efficient estimator'. The book contains accounts of recent work in the theory of probability and statistics and should be included among the modern text-books on these subjects.

It is well printed and the type is easily read. It is not expensive for its size and should be useful both as a general reference book and as a student's text-book.

REGINA C. ELANDT

On the Dynamics of Exploited Fish Populations. By R. J. H. BEVERTON and S. J. HOLT. (Fishery Investigations, series II, volume XIX.) London: H.M.S.O. 1957. Pp. 533. £6. 6s.

It is a pleasure to bring this book to the notice of statisticians and biometricians. The authors, one from the Fisheries Laboratory at Lowestoft and one from the Fisheries Division of the Food and Agriculture Organization at Rome, set out to describe in mathematical language the processes of fishing and seek from this description to predict what the effect of changing methods might be. Part I is concerned with 'Fundamentals of the theory of fishing, illustrated by the analysis of a trawl fishery'. A preliminary model assuming a steady state in a fishery is described and the model is then distorted to allow for other characteristics. In part II extensions of the theory of fishing are discussed. We are introduced to the effect on the fish population of egg production, natural mortality, gear and fishing intensity, and growth and feeding.

Now it is clear that the setting up of mathematical models will involve the estimation of the parameters involved in these models. This estimation is thoroughly discussed in part III, while part IV gives the effect on the model of possible variations in the parameters and attempts prediction. Part IV in fact tries to predict what would be the outcome of attempting to regulate fishing effort and net-mesh size. The authors come to the conclusion that the 'best' results—'best' meaning increased catches, increase in profits and saving in the overheads of men and labour—will be obtained if the fishing effort is restricted to half that of the years before 1939, or less, and if the trawl nets have an enforced 80–90 mm. mesh size.

The statistician will be quite clear that to set up mathematical models is not enough: it is also necessary to verify that the model fits the observed data. The authors are also aware of this necessity. They do not in the course of their book usually give the data on which their models are based. What they do is to refer to the published research work of themselves and others as justification. In the light of the length of the book in even its present state this seems an admirable procedure.

The exposition is lucid and enjoyable. It can be recommended to anyone tired of abstract theory and wanting real applications.

F. N. DAVID

Contribuito allo Studio delle Tavole di Nuzialità. By G. PANIZZON. Padova, Italy: Cedam. 1958. Pp. 143. 1500 lire.

This book is concerned with the construction of tables of the probability of marriage in Italy. The probability of marrying within a year, for a bachelor or spinster of any given age, is termed the 'relative' probability. Since an appreciable proportion of unmarried persons will die, emigrate or immigrate within the year, a correction has to be made for these factors before a valid comparison can be made between marriage rates at different epochs. The author therefore calculates a hypothetical 'absolute probability' of marriage which would result if migration or death could have been prevented. The 'life tables' based on these absolute probabilities are called 'gross' tables. The author shows how to obtain relative and absolute probabilities and net and gross tables from available demographic data, and gives new tables of his own based on the 1951/52 census, comparing these with earlier Italian work, and also with British marriage rates. The general conclusion is that on the average marriage is now postponed, at least for bachelors, to an appreciably later date than in 1900. The modal probability for males has shifted by about 4 years in the last 50 years: in females the difference is rather smaller, and the curve has become appreciably less skew. The book is illustrated by numerous tables and a few graphs. There is also a discussion of abbreviated tables (at, say, 5-year intervals) and rapid methods.

CEDRIC A. B. SMITH

Periodic Regression in Biology and Climatology. By C. I. BLISS. New Haven, Connecticut: Connecticut Agricultural Experimental Station. (Bulletin 615.) 1958. Pp. 55. Gratis.

The author discusses the analysis of periodic data when the length of the cycle is determined independently, where the observations are spaced at even intervals and there is a constant number of replications at each point. Illustrations are given on monthly temperatures, iodine values for butterfat, hourly potential in an elm tree, monthly death rates from pneumonia, monthly incidence of poliomyelitis, hourly incidence of human births, reaction of toads to gonadotrophin and hourly heat exchange of cows. It will be noted that the bulletin is issued free on application.

F. N. DAVID

Queues, Inventories and Maintenance. (Publications in Operations Research, No. 1.)

By PHILIP M. MORSE. New York: John Wiley and Sons Inc.; London: Chapman and Hall Ltd. 1958. Pp. 202. 52s.

The authors of *Finite Queueing Tables* open their preface with the statement that 'there are several text-books and monographs which discuss the theoretical aspects of queueing theory'; nevertheless, this reviewer welcomes Professor Morse's book as the first in English devoted entirely to this subject. His aim is modest: 'The present volume does not pretend to be an exhaustive treatise on queueing theory. Its purpose is primarily expository, to present enough of the concepts, to define some of the terms, and to illustrate a few of the analytic techniques...', and within this framework he is very successful. Apart from a short chapter on transient behaviour (which he dismisses too lightly: 'Transient solutions are not usually of much practical importance') he confines his attention to the steady-state solutions of systems whose inter-arrival and service-time distributions are related to the negative exponential. If he has one major fault it is the making of too general statements. Sometimes he is cautious (p. 140): 'We note that the probabilities are the same no matter what the distribution of replenishment times is (constant, or exponential, or hyper-exponential)'—sometimes he is right (p. 98): 'Eqs. (7.37) and (7.59) show that when the single service channel is exponential, state probabilities P_n form a geometric series no matter what the arrival distribution may chance to be.' But at least once he is wrong (p. 87): 'As with any single-channel, infinite-queue system, the probability that the system is empty is $1 - \rho$.' As the single-channel system denoted (in D. G. Kendall's now standard notation) by $E_2/E_2/1$ one can obtain

$$P_0 = (1 - \rho) \{1 - \frac{1}{2}(1 + \rho - \sqrt{[(1 + \rho)^2 + 4\rho]})\},$$

which is not generally equal to $1 - \rho$.

Technically the book is well produced and the number of misprints is small. There is one point of presentation which could cause confusion to the unwary. On p. 47 the author numbers the phases of an E_i arrival mechanism in the reverse order: on p. 84 he numbers them in the natural order: on p. 108, fig. (8.5) does not correspond to equation (8.6). The notation used for an E_k service mechanism is consistent.

At the end of a discussion of a maintenance model Professor Morse has written: 'All this could have been reasoned out qualitatively without recourse to the equations. By use of the equations, however, we can ensure quantitatively that we use the repair facilities in an optimum manner.' This attitude pervades the book, and although some of his arguments are a little untidy he has written a book which will be useful particularly to the practical man who wants to describe real situations, and useful also to the mathematician coming fresh to queueing theory who wants to know what it is about without being overwhelmed by rigour.

D. M. G. WISHART

Finite Queueing Tables. (Publications in Operations Research, No. 2.) By L. G. PECK and R. N. HAZELWOOD. New York: John Wiley and Sons Inc.; London: Chapman and Hall Ltd. 1958. Pp. 120. 68s.

On internal evidence the book by Professor Morse (reviewed above) appears to have been written before the book under review. It is a pity therefore that Doctors Peck and Hazelwood had not consulted the manuscript of Professor Morse's book before preparing their own: its would greatly have helped the reader (and the two books being of a series they command essentially the same audience) had a uniform notation been adopted. The tables are concerned primarily with problems of maintenance where the population (N) of potential customers is finite. Two quantities, D and F , are tabulated as functions of N , M (the number of repair men), and a parameter X (which is related to the parameter x used by Morse in his discussion of the same problem by the formula $X = 1/(1 + x)$). F is called the efficiency factor (the only inefficiency in the authors' opinion appears to be a machine broken down and waiting for service): D is said to be the probability of a call for service being delayed (it is really one minus the probability that a call received immediate attention—in the case of an infinite population these are the same, but for a finite population we encounter the difficulty that if the entire population is broken down then there are no more calls to be delayed). As D. R. Cox has pointed out to at least one reviewer, the tables appear to be in error. Using formulae 11.12 and the tables in Professor Morse's book this reviewer has derived D for $M = 1$, $X = 0.5$, and $N = 3, 4, 5$; these values agree with D in Peck and Hazelwood's tables for $M = 1$, $X = 0.5$, and $N = 4, 5, 6$ respectively. Conjecture: for N read $N - 1$ throughout.

D. M. G. WISHART

Handbook of Probability and Statistics with Tables. By R. S. BURINGTON and D. C. MAY. U.S.A. and London: McGraw-Hill Book Co. Inc. 1953. Pp. ix + 246 + 72. 46s. 6d.

This book containing 246 pages of expository material and 72 pages of tables is intended as a 'cook-book' for users of statistics rather than an instruction text for the embryo statistician. The idea has been to summarize formulae, definitions, theorems, etc., commonly met with in elementary statistics and probability theory, and to give short tables of the particular functions thus summarized. The topics covered are the binomial, Poisson and normal distributions, regression and time series, tests of significance and confidence intervals, elementary analysis of variance, finite differences and interpolation, and quality control. The tables include the binomial ${}_nC_x p^x q^{n-x}$ for $n = 1(1) \dots 20$, $x = 0(1) \dots n$, $p = 0.05(0.05) \dots 0.50$, the summed binomial for the same set of values, the Poisson distribution $e^{-m} m^x / x!$ for $m = 0.1(0.1) \dots 10$, $11(1) \dots 20$, and the summed Poisson, the normal curve area and ordinates and the first five derivatives, tables of F , z , t , and χ^2 . There are also supporting auxiliary tables. The exposition is adequate without being inspired.

F. N. DAVID

Elementary Matrix Algebra. By FRANZ E. HOHN. New York: The Macmillan Co. 1958. Pp. xi + 305. 52s. 6d.

The author has had considerable experience in presenting the elements of matrix algebra to varied groups of students, including statisticians; the present volume appeared in preliminary editions in 1952 and 1957. The material has been very carefully arranged and graded. The chapter headings are Introduction to Matrix Algebra, Determinants, The Inverse of a Matrix, Rank and Equivalence, Linear Equations and Linear Dependence, Vector Spaces and Linear Transformations, Unitary and Orthogonal Transformations, The Characteristic Equation of a Matrix, and Bilinear, Quadratic, and Hermitian Forms. This final chapter includes a proof of Cochran's theorem. There are appendices on the Σ and Π notations, complex numbers, and isomorphism.

The emphasis throughout is slightly towards abstract algebra—fields, groups, and vector spaces are defined in the course of the work—but these concepts are always carefully introduced and explained in concrete terms. There are many exercises for the student, some of which are marked as comprising results of particular importance.

The author disclaims any pretensions to completeness, and detailed treatment of applications has not been attempted; from the statistical point of view it is unfortunate that he has not found space to mention the multivariate normal distribution, multiple regression, or to give the Helmert and hyperspherical polar transformations explicitly. Nevertheless, this is an excellent book which deserves to become standard.

C. L. MALLOWS

Introduction to Functional Analysis. By A. E. TAYLOR. New York: John Wiley and Sons Inc.; London: Chapman and Hall Ltd. 1958. Pp. 423. 100s.

Professor Taylor avowedly intends that this book shall help the beginner in Functional Analysis who starts with a modest background consisting mainly of a sound Honours Degree course in 'classical Analysis', the additional necessary preparation regarding Linear Spaces and General Topology being provided within the book. He has succeeded in attaining this object, the book being admirably suited to use as a text accompanying early postgraduate training. The subject-matter is clearly presented and well illustrated by examples, the choice of which is not limited to the orthodox (and by now somewhat dated) applications to sequence spaces and Summability Theory. Professor Taylor's personal interest in applications to complex variable problems, for example, has provided a refreshing source of illustrative examples. Other comparatively little known illustrations are provided by discussions of the functional analytic aspects of the Dirichlet Problem and Weyl's 'Projection Method' in Potential Theory. Problems appear in fair abundance and cover a wide range of difficulty. The more difficult questions often accompanied by helpful suggestions, must frequently succeed in stimulating the reader to continue the trend of thought introduced in the text. Most chapters begin with a summary which includes an allocation of rank to, and an indication of the relations between, the major theorems to follow. The

allocation may not always be that which would be awarded by the specialist, but it is in accord with the aim of this book. This same remark applies to the occasional repetition of fundamental results resulting from the natural emphasis on normed linear spaces.

Regarding the specific contents of the book, it is fair to say that, apart from the theory of Banach algebras, the major portion of the basic equipment of a functional analyst is provided. Chapters 1 and 2 set the stage by dealing with Linear Spaces and General Topology respectively, whilst chapter 3 (Topological Linear Spaces) begins the book proper. Some of the topics discussed in chapter 3 receive further attention in chapter 4; in particular, Duality Theory appears in both chapters. Chapter 4 also deals with Representation Theorems for linear functionals and operators, the deeper properties of linear operators (including that happy trio: the Homomorphism, Inversion and Closed Graph Theorems), and boundedness principles for sets of linear operators. Applications of the Riesz Convexity Theorem are also discussed.

Chapters 5 and 6 are both devoted to Spectral Theory, the latter chapter concentrating on the Hilbert space setting. This highly developed field receives a treatment which is both clear and reasonably complete. It includes the Riesz theory for compact operators and the Spectral Decomposition Theorem for bounded self-adjoint and unitary operators; the case of unbounded self-adjoint operators is only briefly treated, details of proofs being omitted and references given.

Chapter 7, entitled 'Integration and Linear Functionals', is perhaps the least satisfactory portion of the book. It provides a guide to existing accounts, without itself providing more than a few of the details. This reviewer suspects that satisfaction will be gained only by the reader who already has some knowledge of the topics under discussion, and for whom the chapter would be largely superfluous anyway. Otherwise full reward will be reserved (perhaps not unjustly) for the keen reader who is prepared to follow up the reading suggested. Abstract measure spaces, Borel measures and Radon measures are all dealt with, albeit rather summarily, as also are bounded, finitely-additive set functions and their integrals. The typography is excellent, and no misprints were found.

One may confidently expect that this book will find a well deserved place on the personal bookshelves of many students and teachers of Functional Analysis, particularly so in the former case if it should be possible to publish a cheaper edition.

R. E. EDWARDS

Tables of Partitions. (Royal Society Mathematical Tables, Vol. 4.) By H. GUPTA, C. E. GWYTHYER and J. C. P. MILLER. Cambridge University Press for the Royal Society. 1958. Pp. xxxix + 132. 63s.

In spite of their title these tables do not list the partitions of numbers but consist essentially of an 87-page table (Table I) of the function $p(n, m)$, the number of partitions of n into at most m parts, for the ranges $n = 1(1)200$, $m = 0(1)\min(n, 100)$ and $n = 201(1)400$, $m = 0(1)50$. The integer-valued entry goes up to nineteen figures at its largest. The excellent and comprehensive introduction reveals how many other allied functions may be easily derived from this, as for instance the number of partitions of n with: greatest part at most m ; exactly m parts; exactly m unequal parts; greatest part exactly m ; exactly m unequal parts (including possibly a zero part); etc. The introduction also gives the exact formulae for $p(n, m)$ for m up to 12 (of Herschel, Cayley, Sylvester & Glaisher) in terms of both circulators and prime circulators. Further, a short discussion of asymptotic formulae is given followed by a description of the method of computation of the tables and a bibliography.

The remainder of the tables consist of: values of $p_2(n, m)$ [Table II, pp. 90-121]; $p_3(n, m)$ [Table III, pp. 122-131] and $p_4(n, 0)$ [Table IV, p. 132], tabulated over a complicated range of values not briefly to be expressed. These functions may be used successively to check and extend the earlier tables in virtue of such relations as

$$p(n, m) = \sum_{r=0}^n (-1)^r p_2(n - rm - \frac{1}{2}r(r+1), r),$$

$$p_2(n, m) = \sum_{r=0}^n (-1)^r p_3(n - rm - \frac{1}{2}r(r+1), r), \quad \text{etc.}$$

These functions are correspondingly larger integers; the function $p_2(n, 0)$, the number of unrestricted partitions of n , being of thirty-two figures for $n = 1000$, for instance. They are generally defined by

$$\sum_{n=0}^{\infty} p_s(n, m) t^n = \{\phi(t, t)\}^{s-1} \sum_{n=0}^{\infty} p(n, m) t^n,$$

where
$$\phi(t, t) = \prod_{r=1}^{\infty} (1-t^r)^{-1}, \quad \sum_{n=0}^{\infty} p(n, m) t^n = \prod_{r=1}^m (1-t^r)^{-1}.$$

The tables are not likely to be of great use to statisticians, who will largely be well satisfied with much shorter earlier tables, but it is very heartening for probabilists to have such extensive tables for reference and use in the rare combinatorial problems which demand them.

D. E. BARTON

Table for the Solution of Cubic Equations. By H. E. SALZER *et al.* U.S.A. and London: McGraw-Hill Book Co. Inc. 1958. Pp. xv + 160. 58s.

The table gives the roots of the general cubic equation

$$ax^3 + bx^2 + cx + d = 0$$

by means of tabulating the roots of $\theta f^3 + f = 1$

as functions f_1, f_2, f_3 of θ . It is left to the reader to compute the argument

$$\theta = \frac{1}{3}(3a^2d - abc + \frac{2}{9}b^3)^2 / (3ac - b^2)^3$$

and convert the f -root back to the x -root by means of

$$x = -\frac{1}{3}b/a - f(3ad - bc + \frac{2}{9}b^3/a) / (3ac - b^2).$$

For $|\theta| \geq 1$ the argument is $1/\theta$; otherwise it is θ and the interval is 0.001. The whole range of θ is thus covered. The entry is to 7 and 8 significant figures, and first and second differences are also given. When two of the roots are complex, as they are outside $(-4/27, 0)$, they are tabulated as $\mathcal{R} + i\mathcal{I}$ (being conjugate). This is in spite of the fact that $-2\mathcal{R}$ equals the real root and the imaginary part is only slightly less easily derived.

The introductory matter tells us that 'linear or quadratic interpolation will yield full accuracy nearly everywhere'. It is suggested that the tables will be commonly useful in what amounts to inverse third-order interpolation, but the reviewer can conceive of no use for them at all which would justify paying the prohibitive price.

D. E. BARTON

Surveys in Applied Mathematics. New York: John Wiley and Sons Inc.; London: Chapman and Hall Ltd. 1958.

Vol. 1. **Elasticity and Plasticity.** By J. N. GOODIER and P. G. HODGE, Jr. Pp. ix + 152. 50s.

Vol. 2. **Dynamics and Nonlinear Mechanics.** By E. LEIMANIS and N. MINORSKY. Pp. xii + 206. 62s.

Vol. 3. **Mathematical Aspects of Subsonic and Transonic Gas Dynamics.** By LIPMAN BERS. Pp. xv + 164. 50s.

Vol. 4. **Some Aspects of Analysis and Probability.** By IRVING KAPLANSKY, MARSHALL HALL JR., EDWIN HEWITT and ROBERT FORTET. Pp. xi + 243. 72s.

Vol. 5. **Numerical Analysis and Partial Differential Equations.** By GEORGE E. FORSYTHE and PAUL C. ROSENBLUM. Pp. x + 204. 60s.

These survey articles have been written as a joint project of the Office of Naval Research and Applied Mechanics Reviews. They are aimed '...not so much at research specialists, actively contributing to the subjects discussed, as...at a broader, mathematically literate audience, looking for contemporary information on the important problems and results in these disciplines'. A major objective is to present recent developments appearing in Russian journals.

The casual reader should be warned that the degree of 'mathematical literacy' assumed is considerable; it is doubtful whether any one person will have the energy, ability or inclination to digest the contents of all these volumes. Certainly your reviewer is not competent to comment on the first three.

The obvious temptation confronting the author of such a survey is to fail to give a balanced picture by placing undue emphasis on his own work; this temptation has not always been avoided. Also the articles will necessarily be out of date almost before they are published; in these volumes there are very few references later than 1956.

The articles most interesting to statisticians *per se* are 'A Survey of Combinatorial Analysis' (Hall) and 'Recent Advances in Probability Theory' (Fortet) in vol. 4, and 'Contemporary State of Numerical Analysis' (Forsythe) in vol. 5. The other articles in vol. 4, namely 'Functional Analysis' (Kaplansky) and 'A Survey of Abstract Harmonic Analysis' (Hewitt) may interest specialists in matters measure-theoretic.

Hall's article on combinatorial analysis (68 pp.) gives only three glancing references to Russian work. The emphasis throughout is entirely different from that of Riordan's recent book (reviewed in the last issue of *Biometrika*); neither Hall nor Riordan give any references to the other's papers. Hall has very little sympathy with methods using generating functions; his first section (Methods of Enumeration) suffers greatly from comparison with Riordan's book. His second part is devoted to the theorem on distinct representatives, with its generalizations and some applications; the discussion in the final part (Existence and Construction of Designs) concentrates on the more elegant (and less practical?) parts of the theory.

Fortet's article (70 pp.) is admittedly selective, omitting much of the theory of stochastic processes. It includes summaries of recent work on axiomatic foundations, central limit theorems (including results on rapidity of convergence and limit theorems for densities), 'general random elements', functionals of random functions, and the Kolmogorov-Smirnov tests. This material is excellent, but will be rather too abstract for the taste of most British statisticians.

Forsythe's article on Numerical Analysis (40 pp.) gives many references to Russian work, and discusses Russian computers (as in 1956). Topics covered in detail are numerical integration, approximation of functions, solution of linear equations, matrix eigenvalue problems, and difference methods for Laplace's equation.

It is difficult to imagine who is going to buy these books. Some large mathematics departments seeking to interest their students may find it worth while; but many potential customers who might have been attracted by one or two articles will be put off by the necessity of buying also uninteresting material which doubles or quadruples the cost.

C. L. MALLOWS

OTHER BOOKS RECEIVED

- Statistics of Extremes.** By E. J. GUMBEL. New York: Columbia University Press. 1958. Pp. 375. \$15.00.
- Statistics: an Introduction.** By DONALD A. S. FRASER. New York: John Wiley and Sons Inc.; London: Chapman and Hall Ltd. 1958. Pp. 398. 54s.
- Information Theory and Statistics.** By SOLOMON KULLBACK. New York: John Wiley and Sons Inc.; London: Chapman and Hall Ltd. 1959. Pp. 394. 100s.
- Statistical Estimates and Transformed Beta-Variables.** By GUNNAR BLOM. New York: John Wiley and Sons Inc.; London: Chapman and Hall Ltd. 1958. Pp. 176. 40s.
- Die Axiomatischen Grundlagen einer Allgemeinen Theorie des Messens.** By J. PFANZAGL. Würzburg, Germany: Physica-Verlag. 1959. Pp. 63. DM.14.
- General Systems.** [Yearbook of the Society for General Systems Research, Vol. 3, 1958.] Michigan, U.S.A.: Braun-Brumfield Inc. Pp. 259. \$7.50.
- Strategy and Market Structure.** By MARTIN SHUBIK. New York: John Wiley and Sons Inc.; London: Chapman and Hall Ltd. 1959. Pp. 387. 64s.
- Statistical Methods in Biology.** By N. T. J. BAILEY. London: English Universities Press. 1959. Pp. 200. 25s.
- Analysis of Straight-Line Data.** By F. S. ACTON. New York: John Wiley and Sons Inc.; London: Chapman and Hall, Ltd. 1959. Pp. 267. 72s.
- Teoria della code.** By G. AVONDO BODINO and F. BRAMBILLA. Milano: Istituto Editoriale Cisalpino. 1959. Pp. 213.
- Elementary Decision Theory.** By HERMAN CHERNOFF and LINCOLN E. MOSES. New York: John Wiley and Sons Inc.; London: Chapman and Hall Ltd. 1959. Pp. xv + 364. 60s.
- Quality Control and Industrial Statistics.** By ACHESON J. DUNCAN. Homewood, Illinois, U.S.A.: Richard D. Irwin Inc. 1959. Pp. 946. \$10.

CORRIGENDA

(1) *Biometrika* (1948), 35, pp. 157-75‘The power function of the test for the difference between two proportions in a 2×2 table.’

By P. B. PATNAIK.

I am indebted to Mr R. L. Turner for pointing out some errors in formulae on pp. 172-3 of this paper:

(a) p. 172, the expressions for h_1 and h_2 should read

$$h_1 = k \sqrt{\left\{ n + \frac{2(1-p_1-p_2) \sqrt{[3n(p_1q_1+p_2q_2)] - 3(p_1q_1+p_2q_2)}}{(p_1+p_2)(2-p_1-p_2)} \right\}},$$

$$h_2 = k \sqrt{\left\{ n + \frac{-2(1-p_1-p_2) \sqrt{[3n(p_1q_1+p_2q_2)] - 3(p_1q_1+p_2q_2)}}{(p_1+p_2)(2-p_1-p_2)} \right\}}.$$

(b) p. 173, in equation (25) for $(1+\theta(1-p_1-p_2))$ read $(1+\theta(1-p_1)-p_2)$, an expression which may also be written as $\theta q_1 + q_2$.

It seems probable that these errors crept in at some stage in the copying of my MS. as the numerical values for the second approximation to the power function given in the examples in Table 3 have been found to be correct.

P. B. PATNAIK

(2) *Biometrika* (1958), 45, pp. 411-20

‘Moment estimators and maximum likelihood.’ By L. R. SHENTON.

(a) p. 415, 2nd line

for $N(\text{var } a_{r+1})^{-1}$ read $(N \text{ var } a_{r+1})^{-1}$.

(b) p. 415, 9th line, equation (18)

for $(\lambda\alpha)^x$ read $\lambda^x \alpha^x$.(3) *Biometrika* (1959), 46, pp. 1-29

‘On the cumulants of renewal processes.’ By W. L. SMITH.

(a) p. 2, 3rd line from bottom

for $-\frac{\mu_2}{\mu_1^2}$ read $-\frac{\mu_2}{2\mu_1^2}$.

(b) p. 18, 14th line,

for $|z_j(\xi)|$ read $|Z_j(\xi)|$.

(c) p. 19, 9th line from bottom

for $s-1$ read $1-s$.





